

STIC-Biotech/ChemLib

CRFE

119557

From: Li, Bao-Qun
Sent: Thursday, April 15, 2004 5:48 PM
To: STIC-Biotech/ChemLib

mej

Please do the sequence homology and interference search for SEQ ID NO 1 of application SN. 10,069,883. Any polypeptide with homology 90% to a polypeptide within the SEQ ID NO: 1. The smaller cut of size is 22 amino acid in SEQ ID NO: 1 or 90% of homology peptide to the amino acid residues from 448-538 of SEQ ID NO: 1. The reference claims are claims 1-11.

BaoQun Li
Art Unit 1648
REM, 3D24.
Tel. 20904.

1-AA-538

mej

RECEIVED
APR 16 2004
STIC-Biotech/ChemLib
(STIC)

3p

L₁₂

①

1-538

3p

L₃

②

③

448-538⑧

20mer

Oligomer search

Point of Contact:

Alexandra Wacławiw

Technical Info. Specialist

Searcher: GM16A02 Tel: 308-4491

Phone: 1471 2-2534

Location: 4-19-04

Date Picked Up: 4-19-04

Date Completed: 15

Searcher Prep/Review: 18

Clerical: Online time:

TYPE OF SEARCH:

NA Sequences: ③

AA Sequences: ③

Structures:

Bibliographic:

Litigation:

Full text:

Patent Family:

Other:

VENDOR/COST (where applic.)

STN:

DIALOG:

Questel/Orbit:

DRLink:

Lexis/Nexis:

Sequence Sys.:

WWW/Internet:

Other (specify): compugen

90 7 1.3 182 15 Q85641 Q85641 moloney mur
 91 7 1.3 183 11 Q64107 Q64107 rattus sp.
 92 7 1.3 187 5 Q967E6 Q967E6 cooperia on
 93 7 1.3 187 15 Q85640 Q85640 murine leuk
 94 7 1.3 187 15 Q83401 Q83401 moloney mur
 95 7 1.3 190 4 Q96H29 Q96H29 homo sapien
 96 7 1.3 190 15 Q88303 Q88303 simian srv-
 97 7 1.3 191 10 Q9LTK0 Q9LTK0 arabidopsis
 98 7 1.3 193 4 Q96DJ2 Q96DJ2 homo sapien
 99 7 1.3 194 5 Q23393 Q23393 caenorhabdi
 100 7 1.3 198 3 Q02754 Q02754 saccharomyc

ALIGNMENTS

RESULT 1

Q9UQFO PRELIMINARY; PRT; 538 AA.
 AC Q9UQFO;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Envelope protein precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=99099005; PubMed=9882319;
 RA Blond-J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
 RA Mandrand-B., Mallet F.;
 RT "Molecular characterization and placental expression of HERV-W, a new
 RT human-endogenous-retrovirus-family.";
 RL J. Virol. 73:1175-1185(1999).
 DR EMBL; AF072506; AAD14546.2;
 DR HSSP; P03385; IMOF.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; ENV_polyprotein; 1.
 KW SIGNAL.
 FT SIGNAL
 FT CHAIN 1 21 POTENTIAL.
 FT CHAIN 22 538 ENVELOPE PROTEIN.
 SQ SEQUENCE 538 AA; 59866 MW; C54648A3C7043970 CRC64;

Query Match 100.0%; Score 538; DB 4; Length 538;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWMORPGNIDAPSVRSLSKGTG 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWMORPGNIDAPSVRSLSKGTG 60
 QY 61 TFTAHTMPNCVHSATLCLMHANTHYWTGMINPSCPGGLGVTCWYTTQTGMSDGGGV 120
 DB 61 TFTAHTMPNCVHSATLCLMHANTHYWTGMINPSCPGGLGVTCWYTTQTGMSDGGGV 120
 QY 121 QDQAREKHKEVLSQTRVHGTSSPYKGLDLSKLETHTRLVSLFNLTITGLHEVSA 180
 DB 121 QDQAREKHKEVLSQTRVHGTSSPYKGLDLSKLETHTRLVSLFNLTITGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPVSIIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 DB 181 QNPTNCWICLPLNFRPVSIIPVPEQWNNFSTEINTTSLVGLPLVSNLEIHTSNLTCVKF 240
 QY 241 SNTYTTNSOCIRWTPPTQIVCLPSGIFPVCGTSAYRCINGSSSEMCFLSLVPPMTIY 300
 DB 241 SNTYTTNSOCIRWTPPTQIVCLPSGIFPVCGTSAYRCINGSSSEMCFLSLVPPMTIY 300
 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYFKLSQELNGDME 360

DB 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYFKLSQELNGDME 360
 QY 361 VADSLVTLQDLNLAADVQLNRRALDLTAERGGLFLGEECCVYVNSQGIIVTEKVK 420
 DB 361 VADSLVTLQDLNLAADVQLNRRALDLTAERGGLFLGEECCVYVNSQGIIVTEKVK 420
 QY 421 IRDRIORRAEELRANTGPGWLLSQMWPWILPFIPLAAIILLLLFGPCIFNLLVNFVSSRI 480
 DB 421 IRDRIORRAEELRANTGPGWLLSQMWPWILPFIPLAAIILLLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLQWEPKQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRNPSAGSS 538
 DB 481 EAVKLQWEPKQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRNPSAGSS 538

RESULT 2

Q9NZG3 PRELIMINARY; PRT; 538 AA.
 AC Q9NZG3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Syncytin precursor (Enverin).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20155476; PubMed=10693809;
 RA Sha M., Lee X., Li X., Veldman G.M., Finnerty H., Racie L.,
 RA LaVallie E., Tang X., Edouard P., Howes S., Keith J.C. Jr.,
 RA McCoy J.M.;
 RT "Syncytin is a captive retroviral envelope protein involved in human
 RT placental morphogenesis.";
 RL Nature 403:785-789(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Alliel P.M., Perin J.-P., Pierig R., Nussbaum J.L., Menard A.,
 RA Rieger F.;
 RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
 RT its env transcripts.";
 RL C.-R. Acad. Sci. III, Sci. Vie 312:857-863(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=21985840; PubMed=11990458;
 RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
 RT "The HERV-W/7q family in the human genome. Potential for protein
 RT expression and gene regulation.";
 RL Cell. Mol. Biol. 48:213-217(2002).
 DR EMBL; AF208161; AAF28334.1;
 DR EMBL; AF513360; AAM47599.1;
 DR HSSP; P03385; IMOF.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; ENV_polyprotein; 1.
 KW SIGNAL.
 FT SIGNAL
 FT CHAIN 1 20 POTENTIAL.
 FT CHAIN 21 538 SYNCYTIN.
 SQ SEQUENCE 538 AA; 59893 MW; B9498850CB34FB78 CRC64;

Query Match 81.2%; Score 437; DB 4; Length 538;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWMORPGNIDAPSVRSLSKGTG 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWMORPGNIDAPSVRSLSKGTG 60

Db 61 TPTAHTMPCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYFTQGMDSGGV 120
QY 121 QDOAREKHVKEVISQLTRVHGTSSPYKGLDLSKLEHETLRLTHRLVSLFNTLTGLHEVSA 180
Db 121 QDOAREKHVKEVISQLTRVHGTSSPYKGLDLSKLEHETLRLTHRLVSLFNTLTGLHEVSA 180
QY 181 QNPTNCWICLPLNFRPYVSIIPVEOWNNFSFTEINTTSVLVGLVSNLEITHSTNLTCVKF 240
Db 181 QNPTNCWICLPLNFRPYVSIIPVEOWNNFSFTEINTTSVLVGLVSNLEITHSTNLTCVKF 240
QY 241 SNTYTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFVPPMTIY 300
Db 241 SNTYTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFVPPMTIY 300
QY 301 TEQDLYSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDMER 360
Db 301 TEQDLYSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDMER 360
QY 361 VADSLVTLQDQNSLAALVQLNRRALDLDLTAERGTCFLGEECCYVYVNSGIVTEKVE 420
Db 361 VADSLVTLQDQNSLAALVQLNRRALDLDLTAERGTCFLGEECCYVYVNSGIVTEKVE 420
QY 421 IRDRIQRRABEELRNTGFWLLSOWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI 480
Db 421 IRDRIQRRABEELRNTGFWLLSOWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQSKTKYRRPLDRPASPRSDVNDIKGTPPEEISAAOPLLRPNAGSS 538
Db 481 EAVKLQMEPKMQSKTKYRRPLDRPASPRSDVNDIKGTPPEEISAAOPLLRPNAGSS 538

RESULT 4

Q8NH7 PRELIMINARY; PRT; 533 AA.
ID Q8NH7; AC Q8NH7; 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Enverin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99052087; PubMed=9835022;
RA Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
RA Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
RT its env transcripts.";
RL C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=21985840; PubMed=11990458;
RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RA "The HERV-W/7q family in the human genome. Potential for protein
RT expression and gene regulation.";
RL Cell. Mol. Biol. 48:213-217(2002).
DR EMBL; AF06835; AAM33413.1;
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR Pfam; PF00429; Env_Polyprotein; 1.
FT NON TER 533
SQ SEQUENCE 533 AA; 59503 MW; 14FB789BCE1C41CD CRC64;

Query Match 80.3%; Score 432; DB 4; Length 533;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 532; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 TPTAHTMPCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYFTQGMDSGGV 120
Db 61 TPTAHTMPCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYFTQGMDSGGV 120
QY 121 QDOAREKHVKEVISQLTRVHGTSSPYKGLDLSKLEHETLRLTHRLVSLFNTLTGLHEVSA 180
Db 121 QDOAREKHVKEVISQLTRVHGTSSPYKGLDLSKLEHETLRLTHRLVSLFNTLTGLHEVSA 180
QY 181 QNPTNCWICLPLNFRPYVSIIPVEOWNNFSFTEINTTSVLVGLVSNLEITHSTNLTCVKF 240
Db 181 QNPTNCWICLPLNFRPYVSIIPVEOWNNFSFTEINTTSVLVGLVSNLEITHSTNLTCVKF 240
QY 241 SNTYTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFVPPMTIY 300
Db 241 SNTYTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFVPPMTIY 300
QY 301 TEQDLYSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDMER 360
Db 301 TEQDLYSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDMER 360
QY 361 VADSLVTLQDQNSLAALVQLNRRALDLDLTAERGTCFLGEECCYVYVNSGIVTEKVE 420
Db 361 VADSLVTLQDQNSLAALVQLNRRALDLDLTAERGTCFLGEECCYVYVNSGIVTEKVE 420
QY 421 IRDRIQRRABEELRNTGFWLLSOWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI 480
Db 421 IRDRIQRRABEELRNTGFWLLSOWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQSKTKYRRPLDRPASPRSDVNDIKGTPPEEISAAOPLLRPNAGSS 538
Db 481 EAVKLQMEPKMQSKTKYRRPLDRPASPRSDVNDIKGTPPEEISAAOPLLRPNAGSS 538

RESULT 3

Q9NRZ2 PRELIMINARY; PRT; 538 AA.
ID Q9NRZ2; AC Q9NRZ2; 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Envelope protein.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20284713; PubMed=10826480;
RA Voisset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
RA Paranhos-Baccala G.;
RT "Chromosomal distribution and coding capacity of the human endogenous
RT retrovirus HERV-W family.";
RL AIDS Res. Hum. Retroviruses 16:731-740(2000).
DR EMBL; AF156963; AF74215.1;
DR HSP; P03385; IMOF.
DR Genew; HGNC:13525; ERVWE1.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR Pfam; PF00429; Env_Polyprotein; 1.
SQ SEQUENCE 538 AA; 59894 MW; 476998A13A2A256F CRC64;

Query Match 81.2%; Score 437; DB 4; Length 538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALPYHIFLTVLLPSFTLTAPPCCRMCTSSSYQBFLEWMPQGNIDAPYSRLSKGTP 60
Db 1 MALPYHIFLTVLLPSFTLTAPPCCRMCTSSSYQBFLEWMPQGNIDAPYSRLSKGTP 60
QY 61 TPTAHTMPCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYFTQGMDSGGV 120

ALIGNMENTS

RESULT 1

US-09-175-928-4
 ; Sequence 4, Application-US/09175928A
 ; Patent No. 6312921

GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: LeVallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Mi, Sha
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
 ; FILE REFERENCE: 6006B.AJ172A
 ; CURRENT APPLICATION NUMBER: US/09/175,928A
 ; CURRENT FILING DATE: 1998-10-20
 ; NUMBER OF SEQ ID NOS: 62
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 538
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-175-928-4

Query Match 81.2%; Score 437; DB 4; Length 538;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | MALPYHIFLFTVLLSFTLTAPPQRCMTSSSPYQEFLLWRMPQGNIDAPYSRLSKGTP | 60 |
| DB | 1 | MALPYHIFLFTVLLSFTLTAPPQRCMTSSSPYQEFLLWRMPQGNIDAPYSRLSKGTP | 60 |
| QY | 61 | TPTAHTMPNCHNSATLCHANTHWGKMINPSCPGGLGVTVCTVYFTQTGMSDGGV | 120 |
| DB | 61 | TPTAHTMPNCHNSATLCHANTHWGKMINPSCPGGLGVTVCTVYFTQTGMSDGGV | 120 |
| QY | 121 | QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLHETLTHTRVLSLFTNTLTGLHEVSA | 180 |
| DB | 121 | QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLHETLTHTRVLSLFTNTLTGLHEVSA | 180 |
| QY | 181 | QNPNCWICLPLNFRPYVSIPEQWNNFSTINTSVLGPLVSNLEITHTSNLTCKVF | 240 |
| DB | 181 | QNPNCWICLPLNFRPYVSIPEQWNNFSTINTSVLGPLVSNLEITHTSNLTCKVF | 240 |
| QY | 241 | SNITYTNSQCIRHWTPPTQIVCLPSGIFPVCGTSAYRCNLGSSSMCFSLFVPPMTIY | 300 |
| DB | 241 | SNITYTNSQCIRHWTPPTQIVCLPSGIFPVCGTSAYRCNLGSSSMCFSLFVPPMTIY | 300 |
| QY | 301 | TEODLYSYVISKPRNKRVPILPVIAGVGLGALGTGIGITTTSTOPYKLSOELNGDWER | 360 |
| DB | 301 | TEODLYSYVISKPRNKRVPILPVIAGVGLGALGTGIGITTTSTOPYKLSOELNGDWER | 360 |
| QY | 361 | VADSLVTLQDLSLAALVQLNRAALDLTAERGTCFLGREGCCYYNQSGIVTEKVK | 420 |
| DB | 361 | VADSLVTLQDLSLAALVQLNRAALDLTAERGTCFLGREGCCYYNQSGIVTEKVK | 420 |
| QY | 421 | IRDRIQRRAEELNTGFWLLSQMFWILFFGLPLAAIILLLLFGPCIFNLVNFVSSRI | 480 |
| DB | 421 | IRDRIQRRAEELNTGFWLLSQMFWILFFGLPLAAIILLLLFGPCIFNLVNFVSSRI | 480 |
| QY | 481 | EAVKIQMEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS | 538 |
| DB | 481 | EAVKIQMEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS | 538 |

RESULT 2

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

326 GAGVLGA 332
24 GAGVLGA 30

RESULT 9

ACPS CHLTR STANDARD; PRT; 119 AA.

30-MAY-2000 (Rel. 39, Last sequence update)
30-MAY-2000 (Rel. 41, Last annotation update)
28-FEB-2003 (Rel. 41, Last sequence update)
DE HOLO-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
GN ACPS OR CT100.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
RN NCBI TaxID=813;
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=990000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;

"Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis."

Science 282:754-759 (1998).
FUNCTION: transfers the 4'-phosphopantetheine moiety from coenzyme
A to a Ser of acyl-carrier protein (By similarity).
-1- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
3',5'-bisphosphate + holo-[acyl-carrier protein].
-1- COFACTOR: Magnesium (By similarity).
-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-1- SIMILARITY: Belongs to the P-Pant transferase superfamily. Acps
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isp-sib.ch/announce/>)
or send an email to license@isp-sib.ch.

EMBL; AE001284; AAC67691.1; -.
PIR; C71556; C71556.
DR HAMAP; MF_00101; -; 1.
DR INTERPRO; IPR008278; 4-PPT transf.
DR INTERPRO; IPR002582; ACPS.
DR INTERPRO; IPR004568; Pantethn_trn.
DR Pfam; PF01648; ACPS; 1.
DR PRODOM; PD004282; ACPS; 1.
DR TIGRFAMs; TIGR00516; acps; 1.
DR TIGRFAMs; TIGR00556; pantethn_trn; 1.
KM TRANSFERASE; Lipid synthetase; Fatty acid biosynthesis; Magnesium;
Complete proteome.
KM METAL
PT METAL
FT METAL
ST METAL
SQ SEQUENCE 119 AA; 13274 MW; DCAE55B623769CF9 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

332 ATGTGIG 338
61 ATGTGIG 67

RESULT 10

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:33:58 ; Search time 45 seconds
(without alignments)

3772.195 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 2871

Sequence: 1 MALPYHIFLFTVLLPSELT.....PPEBISAAQFLRPNSAGSS 538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 100 summaries

Database :

SPTREMBL 25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 2871 | 100.0 | 538 | 4 Q3UQF0 | Q3UQF0 homo sapien |
| 2 | 2868 | 99.9 | 538 | 4 Q3NZG3 | Q3NZG3 homo sapien |
| 3 | 2867 | 99.9 | 538 | 4 Q3NR22 | Q3NR22 homo sapien |
| 4 | 2846 | 99.1 | 533 | 4 Q3NH77 | Q3NH77 homo sapien |
| 5 | 2490 | 86.7 | 542 | 15 Q991W9 | Q991W9 multiple sc |
| 6 | 2194 | 76.4 | 410 | 4 Q95244 | Q95244 homo sapien |
| 7 | 1663 | 57.9 | 319 | 4 Q96L59 | Q96L59 homo sapien |
| 8 | 1657 | 57.7 | 319 | 4 Q96L60 | Q96L60 homo sapien |
| 9 | 1850 | 57.5 | 319 | 4 Q96L61 | Q96L61 homo sapien |
| 10 | 1829 | 42.8 | 242 | 4 Q95245 | Q95245 homo sapien |
| 11 | 929 | 32.4 | 180 | 4 Q96TBS | Q96TBS homo sapien |
| 12 | 857 | 29.9 | 180 | 4 Q96TB1 | Q96TB1 homo sapien |
| 13 | 846 | 29.5 | 180 | 4 Q96TB4 | Q96TB4 homo sapien |
| 14 | 829.5 | 28.9 | 179 | 4 Q96TB3 | Q96TB3 homo sapien |
| 15 | 828 | 28.8 | 180 | 4 Q96TB2 | Q96TB2 homo sapien |
| 16 | 809.5 | 28.2 | 179 | 4 Q96TB8 | Q96TB8 homo sapien |

| | | | | | |
|----|-------|------|-----|-----------|---------------------|
| 17 | 802 | 27.9 | 180 | 4 Q96TB6 | Q96TB6 homo sapien |
| 18 | 778 | 27.1 | 180 | 4 Q96TB7 | Q96TB7 homo sapien |
| 19 | 616.5 | 21.5 | 580 | 15 Q9WP29 | Q9WP29 simian retr |
| 20 | 612 | 21.3 | 564 | 15 Q98654 | Q98654 rd114 retro |
| 21 | 611.5 | 21.3 | 586 | 12 Q9IGU2 | Q9IGU2 fowlpox vir |
| 22 | 602.5 | 21.0 | 580 | 15 Q9WQ03 | Q9WQ03 simian retr |
| 23 | 586 | 20.4 | 620 | 6 Q9GLF7 | Q9GLF7 trichosurus |
| 24 | 567.5 | 19.8 | 618 | 11 Q8BI41 | Q8BI41 mus musculus |
| 25 | 564.5 | 19.7 | 584 | 4 Q9UNM3 | Q9UNM3 homo sapien |
| 26 | 563.5 | 19.6 | 584 | 15 Q9N2K0 | Q9N2K0 herv-h/env6 |
| 27 | 496.5 | 17.3 | 563 | 15 Q9N2J9 | Q9N2J9 herv-h/env6 |
| 28 | 488.5 | 17.0 | 584 | 6 Q8MIB6 | Q8MIB6 pan troglod |
| 29 | 484 | 16.9 | 555 | 15 Q9N2J8 | Q9N2J8 herv-h/env6 |
| 30 | 465.5 | 16.2 | 321 | 15 Q85708 | Q85708 rd114 retro |
| 31 | 426.5 | 14.9 | 638 | 15 Q90LX5 | Q90LX5 porcine end |
| 32 | 425.5 | 14.8 | 638 | 15 Q90LX4 | Q90LX4 porcine end |
| 33 | 425.5 | 14.8 | 641 | 15 Q8Q6Y9 | Q8Q6Y9 porcine end |
| 34 | 424 | 14.8 | 669 | 15 Q9YIS3 | Q9YIS3 murine leuk |
| 35 | 423.5 | 14.8 | 638 | 6 Q62705 | Q62705 sus scrofa |
| 36 | 423.5 | 14.8 | 638 | 15 Q90LX3 | Q90LX3 porcine end |
| 37 | 423 | 14.7 | 665 | 15 Q8UMZ9 | Q8UMZ9 moloney mur |
| 38 | 420.5 | 14.6 | 641 | 15 Q83363 | Q83363 murine leuk |
| 39 | 420 | 14.6 | 673 | 15 Q89816 | Q89816 mus dunni e |
| 40 | 419.5 | 14.6 | 637 | 15 Q85630 | Q85630 friend mink |
| 41 | 419 | 14.6 | 676 | 15 Q7ZGR3 | Q7ZGR3 murine leuk |
| 42 | 417 | 14.5 | 638 | 15 Q9DQ21 | Q9DQ21 murine leuk |
| 43 | 417 | 14.5 | 676 | 15 Q41251 | Q41251 rauscher mu |
| 44 | 416 | 14.5 | 635 | 15 Q9WHJ7 | Q9WHJ7 friend mink |
| 45 | 415.5 | 14.5 | 614 | 15 Q7ZL02 | Q7ZL02 recombinant |
| 46 | 415.5 | 14.5 | 614 | 15 Q7ZL00 | Q7ZL00 recombinant |
| 47 | 414.5 | 14.4 | 614 | 15 Q83380 | Q83380 rat leukeni |
| 48 | 414.5 | 14.4 | 636 | 15 Q85506 | Q85506 murine leuk |
| 49 | 413.5 | 14.4 | 455 | 15 Q90120 | Q90120 mink cell f |
| 50 | 413.5 | 14.4 | 688 | 12 Q01280 | Q01280 retrovirida |
| 51 | 413.5 | 14.4 | 688 | 12 Q01281 | Q01281 retrovirida |
| 52 | 413 | 14.4 | 653 | 15 Q8Q6Z0 | Q8Q6Z0 porcine end |
| 53 | 413 | 14.4 | 676 | 15 Q67456 | Q67456 friend muri |
| 54 | 413 | 14.4 | 676 | 15 Q9DQ23 | Q9DQ23 murine leuk |
| 55 | 412.5 | 14.4 | 642 | 15 Q90R14 | Q90R14 feline leuk |
| 56 | 412.5 | 14.4 | 666 | 15 Q90AE9 | Q90AE9 feline leuk |
| 57 | 412 | 14.4 | 639 | 15 Q8AGX9 | Q8AGX9 pythion molu |
| 58 | 412 | 14.4 | 639 | 15 Q8AGX8 | Q8AGX8 pythion molu |
| 59 | 412 | 14.4 | 665 | 15 Q67649 | Q67649 gibbon ape |
| 60 | 412 | 14.4 | 676 | 15 Q85502 | Q85502 friend muri |
| 61 | 411.5 | 14.3 | 638 | 6 Q62707 | Q62707 sus scrofa |
| 62 | 410 | 14.3 | 676 | 15 Q39737 | Q39737 friend muri |
| 63 | 410 | 14.3 | 676 | 15 Q12374 | Q12374 murine leuk |
| 64 | 409.5 | 14.3 | 653 | 15 Q8Q6Z1 | Q8Q6Z1 porcine end |
| 65 | 409.5 | 14.3 | 659 | 15 Q8Q6Y6 | Q8Q6Y6 porcine end |
| 66 | 409.5 | 14.3 | 669 | 15 Q7SVK6 | Q7SVK6 murine leuk |
| 67 | 409 | 14.2 | 659 | 6 Q9TTC0 | Q9TTC0 phascollarct |
| 68 | 409 | 14.2 | 660 | 15 Q8JEM7 | Q8JEM7 porcine end |
| 69 | 408.5 | 14.2 | 590 | 11 Q81876 | Q81876 mus musculus |
| 70 | 408.5 | 14.2 | 620 | 15 Q952L2 | Q952L2 mus cervico |
| 71 | 408.5 | 14.2 | 659 | 15 Q9Q9X3 | Q9Q9X3 porcine end |
| 72 | 408.5 | 14.2 | 664 | 15 Q90RL3 | Q90RL3 murine leuk |
| 73 | 408.5 | 14.2 | 669 | 15 Q83382 | Q83382 murine leuk |
| 74 | 407.5 | 14.2 | 642 | 15 Q85518 | Q85518 feline leuk |
| 75 | 407.5 | 14.2 | 642 | 15 Q89812 | Q89812 feline leuk |
| 76 | 407.5 | 14.2 | 660 | 15 Q8J4V7 | Q8J4V7 porcine end |
| 77 | 407.5 | 14.2 | 669 | 11 Q80589 | Q80589 mus musculus |
| 78 | 407.5 | 14.2 | 671 | 11 Q91Y75 | Q91Y75 mus musculus |
| 79 | 407.5 | 14.2 | 682 | 15 Q83399 | Q83399 murine leuk |
| 80 | 406.5 | 14.2 | 642 | 15 Q85522 | Q85522 feline leuk |
| 81 | 406.5 | 14.2 | 669 | 11 Q04586 | Q04586 mus musculus |
| 82 | 406 | 14.1 | 636 | 15 Q83365 | Q83365 moloney mur |
| 83 | 406 | 14.1 | 661 | 15 Q3IUF7 | Q3IUF7 porcine end |
| 84 | 406 | 14.1 | 670 | 15 Q9YWM0 | Q9YWM0 gibbon ape |
| 85 | 405.5 | 14.1 | 652 | 15 Q83364 | Q83364 moloney mur |
| 86 | 405.5 | 14.1 | 654 | 15 Q89683 | Q89683 murine leuk |
| 87 | 404.5 | 14.1 | 640 | 15 Q9WHV5 | Q9WHV5 murine leuk |
| 88 | 404.5 | 14.1 | 641 | 11 Q8K080 | Q8K080 mus musculus |
| 89 | 404.5 | 14.1 | 654 | 15 Q7ZJT7 | Q7ZJT7 amphotropic |

90 404.5 14.1 659 15 Q8Q6Y8 Q8Q6Y8 porcine end
 91 404.5 14.1 660 15 Q8Q6U6 Q8Q6U6 porcine end
 92 404.5 14.1 2378 15 Q9Q1X4 Q9Q1X4 porcine end
 93 403.5 14.1 641 11 Q811N4 Q811N4 mus musculus
 94 403.5 14.1 642 15 Q855Z5 Q855Z5 feline leuk
 95 403.5 14.1 659 15 Q8Q6V7 Q8Q6V7 porcine end
 96 403.5 14.1 660 15 Q811Z2 Q811Z2 porcine end
 97 403.5 14.1 669 11 Q819I9 Q819I9 mus musculus
 98 403.5 14.1 669 15 Q8Q2Q4 Q8Q2Q4 murine leuk
 99 403 14.0 645 11 Q811M8 Q811M8 mus musculus
 100 402.5 14.0 361 15 Q9YY58 Q9YY58 mouse endog

ALIGNMENTS

RESULT 1

Q9UQF0
 ID Q9UQF0 PRELIMINARY; PRT; 538 AA.
 AC Q9UQF0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Envelope protein precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=9909005; PubMed=9882319;
 RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
 RA Mandrand B., Mallet F.;
 RT "Molecular characterization and placental expression of HERV-W, a new
 RT human endogenous retrovirus family.";
 RL J. Virol. 73:1175-1185(1999).
 DR EMBL; AF072506; AAD14546.2; -;
 DR HSSP; P03385; IMOF.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0005198; P:Structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Signal.
 FT SIGNAL
 FT CHAIN
 SQ SEQUENCE 538 AA; 59866 MW; C54648A3C7043870 CRC64;

Query Match 100.0%; Score 2871; DB 4; Length 538;
 Best Local Similarity 100.0%; Pred. No. 2.5e-247; Mismatches 0; Indels 0; Gaps 0;
 Matches 538; Conservative 0;
 QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
 QY 61 TETATHPRNCHSATLCMHANTHYWTGKMINSPCGGLGVTCWYFTGTGMSDGGV 120
 DB 61 TETATHPRNCHSATLCMHANTHYWTGKMINSPCGGLGVTCWYFTGTGMSDGGV 120
 QY 121 QOAREKHVKEVISQITRVHGTSPPYKGLDLSKHLHETRLTRVLSLFTNLTGLHEVSA 180
 DB 121 QOAREKHVKEVISQITRVHGTSPPYKGLDLSKHLHETRLTRVLSLFTNLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPYVISIPVPEOWNPSTEINTSVLGVPLVSNLEITHTSNLTCKVF 240
 DB 181 QNPTNCWICLPLNFRPYVISIPVPEOWNPSTEINTSVLGVPLVSNLEITHTSNLTCKVF 240
 QY 241 SNTTITNSQCIIRWTPPTQIVCLPSGIFFFVCGTSAYRCLNGSSSESMCFSLFLVPPMTIY 300
 DB 241 SNTTITNSQCIIRWTPPTQIVCLPSGIFFFVCGTSAYRCLNGSSSESMCFSLFLVPPMTIY 300
 QY 301 TEQDLYSYVSKRNRKRVLPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDME 360

DB 301 TEQDLYSYVSKRNRKRVLPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDME 360
 QY 361 VADSLVTLOQLNSLAAYVQLNRALDLTAERGGTCLFLGEBCCYVNVQSGIVTEKVK 420
 DB 361 VADSLVTLOQLNSLAAYVQLNRALDLTAERGGTCLFLGEBCCYVNVQSGIVTEKVK 420
 QY 421 IRDRIQRARELNTGFWGLLSQMPWIIFFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 DB 421 IRDRIQRARELNTGFWGLLSQMPWIIFFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLQNEPQWQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538
 DB 481 EAVKLQNEPQWQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538

RESULT 2

Q9NZG3
 ID Q9NZG3 PRELIMINARY; PRT; 538 AA.
 AC Q9NZG3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Syncytin precursor (Enverin).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20155476; PubMed=10693809;
 RA Sha M., Lee X., Li X., Veldman G.M., Finnerty H., Racie L.,
 RA Lavallie E., Tang X., Edouard P., Howes S., Keith J.C. Jr.;
 RA McCoy J.M.;
 RT "Syncytin is a captive retroviral envelope protein involved in human
 RT placental morphogenesis.";
 RL Nature 403:785-789(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
 RA Rieger F.;
 RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
 RT its env transcripts.";
 RL C. R. Acad. Sci. III, Sci. Vie 312:857-863(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=21985840; PubMed=11990458;
 RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
 RT "The HERV-W/7q family in the human genome. Potential for protein
 RT expression and gene regulation.";
 RL Cell. Mol. Biol. 48:213-217(2002).
 DR EMBL; AF208161; AAF28334.1; -;
 DR EMBL; AF513360; AAM47599.1; -;
 DR HSSP; P03385; IMOF.
 DR GO; GO:0019028; C:Viral capsid; IEA.
 DR GO; GO:0005198; P:Structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Signal.
 FT SIGNAL
 FT CHAIN
 SQ SEQUENCE 538 AA; 59893 MW; B9498850CB34PB78 CRC64;

Query Match 99.9%; Score 2868; DB 4; Length 538;
 Best Local Similarity 99.8%; Pred. No. 4.6e-247;
 Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60

QY 61 TFTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYQTQMSDGGGV 120
 Db 61 TFTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYQTQMSDGGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETRLHTRVLSLNTLTGLHEVSA 180
 Db 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETRLHTRVLSLNTLTGLHEVSA 180
 QY 181 QNPNCWCICLPNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEIHTHTSNLTCVKF 240
 Db 181 QNPNCWCICLPNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEIHTHTSNLTCVKF 240
 QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCNLNGSSCMCFSLFVPPMTIY 300
 Db 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCNLNGSSCMCFSLFVPPMTIY 300
 QY 301 TEQDLYSVYVSKPRNKRPVILPFVIGAGVLCALGTGIGITTSQFYKLSQELNGDMER 360
 Db 301 TEQDLYSVYVSKPRNKRPVILPFVIGAGVLCALGTGIGITTSQFYKLSQELNGDMER 360
 QY 361 VADSLVTLQDQNSLAAYVQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 Db 361 VADSLVTLQDQNSLAAYVQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 QY 421 IRDRIQRAAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 Db 421 IRDRIQRAAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538
 Db 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538

RESULT 3

Q9NRZ2 ID Q9NRZ2 PRELIMINARY; PRT; 538 AA.
 AC 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DE Envelope protein.
 GN ENV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20284713; PubMed=10826480;
 RA Volset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
 RA Paranhos-Baccala G.
 RT "Chromosomal distribution and coding capacity of the human endogenous
 RT retrovirus HERV-W family."
 RL AIDS Res. Hum. Retroviruses 16:731-740(2000).
 DR EMBL; AF156963; AAF74215.1; -.
 DR HSSP; P03385; 1MOF.
 DR Genew; HGNC:13525; BRVWE1.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; ENV_polyprotein; 1.
 DR SEQUENCE 538 AA; 59894 MW; 476998A13A2A256F CRC64;

Query Match 99.9%; Score 2867; DB 4; Length 538;
 Best Local Similarity 99.8%; Pred. No. 5.6e-247;
 Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHFLFTVLPSFTLTAPPCCRCMTSSSPYQEFILWQRGNIDAPYSLSKGTG 60
 Db 1 MALPYHFLFTVLPSFTLTAPPCCRCMTSSSPYQEFILWQRGNIDAPYSLSKGTG 60
 QY 61 TFTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYQTQMSDGGGV 120

Db 61 TFTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVCWTFYQTQMSDGGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETRLHTRVLSLNTLTGLHEVSA 180
 Db 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETRLHTRVLSLNTLTGLHEVSA 180
 QY 181 QNPNCWCICLPNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEIHTHTSNLTCVKF 240
 Db 181 QNPNCWCICLPNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEIHTHTSNLTCVKF 240
 QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCNLNGSSCMCFSLFVPPMTIY 300
 Db 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCNLNGSSCMCFSLFVPPMTIY 300
 QY 301 TEQDLYSVYVSKPRNKRPVILPFVIGAGVLCALGTGIGITTSQFYKLSQELNGDMER 360
 Db 301 TEQDLYSVYVSKPRNKRPVILPFVIGAGVLCALGTGIGITTSQFYKLSQELNGDMER 360
 QY 361 VADSLVTLQDQNSLAAYVQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 Db 361 VADSLVTLQDQNSLAAYVQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 QY 421 IRDRIQRAAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 Db 421 IRDRIQRAAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538
 Db 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNASGSS 538

RESULT 4

Q8NH7 ID Q8NH7 PRELIMINARY; PRT; 533 AA.
 AC Q8NH7;
 DT 01-OCT-2002 (TremBLrel. 22, Created)
 DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)
 DE Enverin (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Placenta;
 RC MEDLINE=99052087; PubMed=9835022;
 RA Alliel F.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
 RA Rieger F.;
 RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
 RT its env transcripts."
 RL C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=21985840; PubMed=11990458;
 RA Alliel F.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
 RT "The HERV-W/7q family in the human genome. Potential for protein
 RT expression and gene regulation."
 RL Cell. Mol. Biol. 48:213-217(2002).
 DR EMBL; AF506835; AM33413.1; -.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; ENV_polyprotein; 1.
 DR NON TER 533 533
 SQ SEQUENCE 533 AA; 59503 MW; 14FB789BCE1C41CD CRC64;

Query Match 99.1%; Score 2846; DB 4; Length 533;
 Best Local Similarity 99.8%; Pred. No. 4.1e-245;
 Matches 532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLEWRMORPGNIDAPSYSLSKGTP 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLEWRMORPGNIDAPSYSLSKGTP 60
QY 61 TFTAHTMPRNCVHSATLCHWANTHYWTGKMINPSCPGGLGVTVCTWTFYQTGMSDGGGV 120
DB 61 TFTAHTMPRNCVHSATLCHWANTHYWTGKMINPSCPGGLGVTVCTWTFYQTGMSDGGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETLRTHTRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETLRTHTRLVSLFNTLTGLHEVSA 180
QY 181 QNPNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTVCVF 240
DB 181 QNPNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTVCVF 240
QY 241 SNTTYTNSOCIRWTPPTQIVCLPSGIFVCGTSAIRCINGSSSMCFSLFVPPMTIY 300
DB 241 SNTTYTNSOCIRWTPPTQIVCLPSGIFVCGTSAIRCINGSSSMCFSLFVPPMTIY 300
QY 301 TEODLYSYVSKPRNKRVPILPFGVAGVLGALGTGIGITSTOPYKLSOELNGDMER 360
DB 301 TEODLYSYVSKPRNKRVPILPFGVAGVLGALGTGIGITSTOPYKLSOELNGDMER 360
QY 361 VADSLVTLQDQLNSLAAVLQNRALDLLTAERGTCFLGEECCYYVNSGIVTEKVE 420
DB 361 VADSLVTLQDQLNSLAAVLQNRALDLLTAERGTCFLGEECCYYVNSGIVTEKVE 420
QY 421 IRDRIOORAEELNTPGWLSSOMWPIPLFGLPLAAIILLFGPCIFNLVNFVSSRI 480
DB 421 IRDRIOORAEELNTPGWLSSOMWPIPLFGLPLAAIILLFGPCIFNLVNFVSSRI 480
QY 481 EAVKLQWEPKQSKTKIYRPLDRPASRSDVNDIKGTPPEEISAQPLLRN 533
DB 481 EAVKLQWEPKQSKTKIYRPLDRPASRSDVNDIKGTPPEEISAQPLLRN 533

RESULT 5

Q991W9 PRELIMINARY; PRT; 542 AA.
ID Q991W9
AC Q991W9
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Recombinant envelope protein.
GN ENV.
OS Multiple sclerosis associated retrovirus element.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=89382;
RN [1]
RP SEQUENCE FROM N.A.
RA Perron H., Jouvin-Marche B., Michel M., Ounanian-Paraz A., Camelo S.,
RA Dumon A., Jolivet-Raynaud C., Marcel F., Souillet Y., Borel E.,
RA Gebuhrer L.;
RA "Multiple sclerosis retrovirus-induced polyclonal Vb6 T-cell receptor
RT activation as a trigger of abnormal immune response";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF311500; AAX18189.1;
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0019031; C:Viral envelope; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Envelope protein.
SQ SEQUENCE 542 AA; 60625 MW; AAFYA8322E03C9EE CRC64;

Query Match 86.7%; Score 2490; DB 15; Length 542;
Best Local Similarity 87.3%; Pred. No. 2.5e-213;
Matches 473; Conservative 17; Mismatches 48; Indels 4; Gaps 1;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLEWRMORPGNIDAPSYSLSKGTP 60

DB 1 MALPYHIFLFTVLLPPPALTAPPCCCTSSSPYQEFLEWRTRLPNGNDAPSYSLSKGNS 60
QY 61 TFTAHTMPRNCVHSATLCHWANTHYWTGKMINPSCPGGLGVTVCTWTFYQTGMSDGGGV 120
DB 61 TFTAHTMPRNCVHSATLCHWANTHYWTGKMINPSCPGGLGVTVCTWTFYQTGMSDGGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETLRTHTRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKHLHETLRTHTRLVSLFNTLTGLHEVSA 180
QY 181 QNPNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTVCVF 240
DB 181 QNPNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTVCVF 240
QY 241 SNTTYTNSOCIRWTPPTQIVCLPSGIFVCGTSAIRCINGSSSMCFSLFVPPMTIY 300
DB 241 SNTTYTNSOCIRWTPPTQIVCLPSGIFVCGTSAIRCINGSSSMCFSLFVPPMTIY 300
QY 301 TEODLYSYVSKPRNKRVPILPFGVAGVLGALGTGIGITSTOPYKLSOELNGDMER 360
DB 301 TEODLYSYVSKPRNKRVPILPFGVAGVLGALGTGIGITSTOPYKLSOELNGDMER 360
QY 361 VADSLVTLQDQLNSLAAVLQNRALDLLTAERGTCFLGEECCYYVNSGIVTEKVE 420
DB 361 VADSLVTLQDQLNSLAAVLQNRALDLLTAERGTCFLGEECCYYVNSGIVTEKVE 420
QY 421 IRDRIOORAEELNTPGWLSSOMWPIPLFGLPLAAIILLFGPCIFNLVNFVSSRI 480
DB 421 IRDRIOORAEELNTPGWLSSOMWPIPLFGLPLAAIILLFGPCIFNLVNFVSSRI 480
QY 481 EAVKLQWEPKQSKTKIYRPLDRPASRSDVNDIKGTPPEEISAQPLLRN 536
DB 481 EAVKLQWEPKQSKTKIYRPLDRPASRSDVNDIKGTPPEEISAQPLLRN 536
QY 537 SS 538
DB 541 SS 542

RESULT 6

Q95244 PRELIMINARY; PRT; 410 AA.
ID Q95244
AC Q95244
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99099005; PubMed=9883319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Maillet F.;
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family";
RL J. Virol. 73:1175-1185 (1999).
DR EMBL; AF072505; AADI4545.1; .
DR HSP; P03385; IMOF.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR GO; GO:0006949; P:syncytium formation; TAS.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 410
SQ SEQUENCE 410 AA; 45442 MW; 637F1ECB35DE1587 CRC64;

Query Match 76.4%; Score 2194; DB 4; Length 410;
Best Local Similarity 99.0%; Pred. No. 4.5e-187;
Matches 406; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DR EMBL; AY050299; AAL11493.1; -.
FT NON TER 319 319
SQ SEQUENCE 319 AA; 35655 MW; 7D3397CCCAAD8999 CRC64;
Query Match 57.5%; Score 1650; DB 4; Length 319;
Best Local Similarity 94.7%; Pred. No. 1.1e-138;
Matches 302; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 MALPYHIFLFTVLPLSTLAPPCCRCMTSSSYQVEFLWMQPGNIDAPSYSLSGTTP 60
DB 1 MALPYHIFLFTVLPLSTLAPPCCRCMTSSSYQVEFLWMQPGNIDAPSYSLSGTTP 60
QY 1 TETAHTMPNCHXSATLCHXANTHYWTGKWNIPSCPGGLGVTVCTYFTQTCMSGGGV 120
DB 1 TETAHTMPNCHXSATLCHXANTHYWTGKWNIPSCPGGLGVTVCTYFTQTCMSGGGV 120
QY 121 QDQAREKHKEVISQLTRVHGTSSPYKGLDLKLHETLRTHTRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHKEVISQLTRVHGTSSPYKGLDLKLHETLRTHTRLVSLFNTLTGLHEVSA 180
QY 181 QNPTNCHICLPLFRPVVSIPEQWNNFSTEINTSVLGPVLSNLEITHSNTLCVKF 240
DB 181 QNPTNCHICLPLFRPVVSIPEQWNNFSTEINTSVLGPVLSNLEITHSNTLCVKF 240
QY 241 SNTTYTNSCIRWVTPTQIVCLPSGIFVCGTSAVRCCLNGSSESCFSLFLVPPMTIY 300
DB 241 SNTTYTNSCIRWVTPTQIVCLPSGIFVCGTSAVRCCLNGSSESCFSLFLVPPMTIY 300
QY 301 TEQDLYSYVSKPRNKRPV 319
DB 301 TEQDLYSYVSKPRNKRPV 319
RESULT 10
ID O95245 PRELIMINARY; PRT; 242 AA.
AC O95245
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99099005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.;
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072508; AAD14548.1; -.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 1
SQ SEQUENCE 242 AA; 26857 MW; A03512693D346FB9 CRC64;
Query Match 42.8%; Score 1229; DB 4; Length 242;
Best Local Similarity 99.2%; Pred. No. 2.9e-101;
Matches 240; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 297 MTIYTEQDLYSYVSKPRNKRPVLPVIGAGVGLGALTGGIGTITSTQFYKLSQELNG 356
DB 1 MAIYTEQDLYSYVSKPRNKRPVLPVIGAGVGLGALTGGIGTITSTQFYKLSQELNG 60
QY 357 DMERVADSLVTLQDQLNSLAAVLQNRALDILTAERGTCFLGEECCYYNQSGIVTE 416

DB 61 DMERVADSLVTLQDQLNSLAAVLQNRALDILTAERGTCFLGEECCYYNQSGIVTE 120
QY 417 KVKEIRDRIQRAAEELRNTGWLISQWMPWILFLGPLAAIILLLFLGPCIIFNLLNVFV 476
DB 121 KVKEIRDRIQRAAEELRNTGWLISQWMPWILFLGPLAAIILLLFLGPCIIFNLLNVFV 180
QY 477 SSRIBAVKLQWPKMQSKTKYRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAG 536
DB 181 SSRIBAVKLQWPKMQSKTKYRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAG 240
QY 537 SS 538
DB 241 SS 242
RESULT 11
ID Q96TB5 PRELIMINARY; PRT; 180 AA.
AC Q96TB5
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268769; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
DR EMBL; AB051004; BAB47558.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 1
FT NON TER 180
SQ SEQUENCE 180 AA; 19648 MW; 8D2F33813F16090B CRC64;
Query Match 32.4%; Score 929; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.2e-74;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 207 NNFSTEINTSVLGPVLSNLEITHSNTLCVKFSNTTYTNSCIRWVTPTQIVCLPS 266
DB 1 NNFSTEINTSVLGPVLSNLEITHSNTLCVKFSNTTYTNSCIRWVTPTQIVCLPS 60
QY 267 GIFFVCGTSAVRCCLNGSSESCFSLFLVPPMTIYTEQDLYSYVSKPRNKRPVLPFVIG 326
DB 61 GIFFVCGTSAVRCCLNGSSESCFSLFLVPPMTIYTEQDLYSYVSKPRNKRPVLPFVIG 120
QY 327 AGVLGALTGGIGTITSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVLQNRAL 386
DB 121 AGVLGALTGGIGTITSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVLQNRAL 180
RESULT 12
ID Q96TB1 PRELIMINARY; PRT; 180 AA.
AC Q96TB1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21268768; PubMed=11375061;
RX Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RL identification, and phylogeny.";
DR EMBL; AB051010; BAB47562.1; -
DR GO; GO:0019028; C: viral capsid; IEA.
DR InterPro; IPR002050; F: structural molecule activity; IEA.
DR Pfam; PF00429; ENV_polyprotein; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 19759 MW; C4D3C9F72388A0C3 CRC64;

Query Match 29.9%; Score 857; DB 4; Length 180;
Best Local Similarity 91.7%; Pred. No. 3.1e-68;
Matches 165; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTYTTNSQCIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTIDTTSQCIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 326
Db 61 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 120

QY 327 AGVLGALGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 386
Db 121 AGVLGRLGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 180

RESULT 13
Q96TB4
ID Q96TB4 PRELIMINARY; PRT; 180 AA.
AC Q96TB4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21268768; PubMed=11375061;
RX Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RL identification, and phylogeny.";
DR EMBL; AB051007; BAB47559.1; -
DR GO; GO:0019028; C: viral capsid; IEA.
DR InterPro; IPR002050; F: structural molecule activity; IEA.
DR Pfam; PF00429; ENV_polyprotein; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 20048 MW; F2B7B2FAF7CE40C3 CRC64;

Query Match 29.5%; Score 846; DB 4; Length 180;
Best Local Similarity 91.7%; Pred. No. 3e-67;
Matches 165; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTYTTNSQCIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTIDTTSQCIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 326
Db 61 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 120
```

```
Db 61 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 120
QY 327 AGVLGALGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 386
Db 121 AGVLGRLGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 180

RESULT 14
Q96TB3
ID Q96TB3 PRELIMINARY; PRT; 179 AA.
AC Q96TB3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21268768; PubMed=11375061;
RX Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RL identification, and phylogeny.";
DR EMBL; AB051008; BAB47560.1; -
DR GO; GO:0019028; C: viral capsid; IEA.
DR InterPro; IPR002050; F: structural molecule activity; IEA.
DR Pfam; PF00429; ENV_polyprotein; 1.
FT NON_TER 1
FT NON_TER 179
SQ SEQUENCE 179 AA; 19471 MW; 70DF3B585B8E667 CRC64;

Query Match 28.9%; Score 829.5; DB 4; Length 179;
Best Local Similarity 90.6%; Pred. No. 8.7e-66;
Matches 163; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

QY 207 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTYTTNSQCIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTIDTTSQCIRWVTPPTQIVCLPS 59

QY 267 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 326
Db 60 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 119

QY 327 AGVLGALGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 386
Db 120 AGVLGRLGTGGIGITTTQFYFKLSQELNGDMERVADSLVTLQDQNSLAAVVLQNRRL 179

RESULT 15
Q96TB2
ID Q96TB2 PRELIMINARY; PRT; 180 AA.
AC Q96TB2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21268768; PubMed=11375061;
RX Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RL identification, and phylogeny.";
DR EMBL; AB051009; BAB47561.1; -
DR GO; GO:0019028; C: viral capsid; IEA.
DR InterPro; IPR002050; F: structural molecule activity; IEA.
DR Pfam; PF00429; ENV_polyprotein; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 20048 MW; F2B7B2FAF7CE40C3 CRC64;

Query Match 29.5%; Score 846; DB 4; Length 180;
Best Local Similarity 91.7%; Pred. No. 3e-67;
Matches 165; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTYTTNSQCIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLGPVLSNLEIHTSNLCVCXFSNTIDTTSQCIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 326
Db 61 GIFFVCGTSAYHCLNGSSSFCFLSFLVPPMTIYTEODLYSYVLSKPKRKRVPILPFVIR 120
```


DR EMBL; AB051009; BAB47561.1; -;
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 180
FT NON_TER 180
SQ SEQUENCE 180 AA; 19709 MW; 4C8F196C9C788497 CRC64;

Query Match 28.8%; Score 828; DB 4; Length 180;
Best Local Similarity 88.3%; Pred. No. 1.2e-65;
Matches 159; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYRCCLNGSSSMCLFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 326
Db 61 GIFFVCGTSAYRCFKSSSICFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 120

QY 327 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 386
Db 121 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 180

RESULT 16
Q96TB8 PRELIMINARY; PRT; 179 AA.
ID Q96TB8
AC Q96TB8
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB051009; BAB47555.1; -;
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 179
FT NON_TER 179
SQ SEQUENCE 179 AA; 19228 MW; 3A9FACBABA6376E74 CRC64;

Query Match 28.2%; Score 809.5; DB 4; Length 179;
Best Local Similarity 89.4%; Pred. No. 5.3e-64;
Matches 161; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

QY 207 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYRCCLNGSSSMCLFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 326
Db 61 GIFFVCGTSAYRCCLNGSSSMCLFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 120

QY 327 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 386
Db 121 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 179

RESULT 17

Q96TB6 PRELIMINARY; PRT; 180 AA.
ID Q96TB6
AC Q96TB6
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB051009; BAB47557.1; -;
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 180
FT NON_TER 180
SQ SEQUENCE 180 AA; 19656 MW; EF8DFCDB4CAEAF3D CRC64;

Query Match 27.9%; Score 802; DB 4; Length 180;
Best Local Similarity 88.3%; Pred. No. 2.5e-63;
Matches 159; Conservative 3; Mismatches 18; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 266
Db 1 NNFSTEINTSVLVGLVSNLEIHTSNTLCVKFSNTIYTSQCIIRWVTPPTQIVCLPS 60

QY 267 GIFFVCGTSAYRCCLNGSSSMCLFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 326
Db 61 GIFFVCGTSAYRCCLNGSSSMCLFLVPPMTIYTEQDLYSVVISKPRNKRVPILPFFVIG 120

QY 327 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 386
Db 121 AGVLGALGTGIGITSTQFYKLSQELNGDMERVADSLVTLDQQLNSLAAVVQNRRAL 180

RESULT 18
Q96TB7 PRELIMINARY; PRT; 180 AA.
ID Q96TB7
AC Q96TB7
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB050999; BAB47556.1; -;
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 180
FT NON_TER 180
SQ SEQUENCE 180 AA; 19984 MW; 3414ECDD74A060BA CRC64;

Best Local Similarity 29.2%; Pred. No. 8.4e-45;
Matches 155; Conservative 96; Mismatches 190; Indels 89; Gaps 16;

QY 17 FTLAPP-----PCRMTSSPYQEFLLRMQRFCNIDAPYSLSKG-----TPTFT 63
DB 54 YVITAPTLYLVATVSCSSHTAYQPSDSLKW-----CVSNPTLANGENIGNCPCQTFK 105
QY 64 AHTMPRCNYSATLCMHANTHYTG-----XMINPSCGGL 100
DB 106 ESVH---SSCYTYQCFNGKTYTIALASNRAPITGTSNVTVLGNHLLSAGCTGTV 163
QY 101 GVTWCWYFTQTGSDGGVQDQAREKHVKEVISQLTRVHGTSSP---YKGLDLSKLHET 157
DB 164 GCHICWPKAPVHISDGGGPDQKAREIAVQK---RLBEIHSRLEPRLYHPLALPKARGK 220
QY 158 LRTHRLVSLNLTGLHEVSAQNPTNCWICLPLNFRPVYISIPVPEQNNFSTEINTTS 217
DB 221 EKIDAQTNLTATVSLNKNPNLANECWICLP-SGNP-VPLAIPS---NDSFGLSLS 276
QY 218 VLVGP-LVSNLEITHSNLTC-----VKFNTTYTNSQCIKRWTPPTQ 260
DB 277 PIIPPLVQPLEFINLASCLYSPSQNSFDVGLVEFTNCSTLNI-----SH 327
QY 261 IVCLPSGIFVCGTS-AVRLNGSSESMCLFSLVPPMTIY-----TEQDIYSVISKPR 314
DB 328 SLCAFNSSVFCGNNKAYTILPTNWTGTCVLATLPLDIDIVGDPAPVPAIDHYLHRAR 387
QY 315 NKRVPLPFGVIGALGTGIGGITTSTQFYKLSOELNGDMRVADSLVT-ODQLNS 374
DB 388 -RAVOFIELLVGLGITTAVSTGTGLGYSIQYKLSQLSDVQAISSTIQDLOQVDS 446
QY 375 LAAVLQNRALDLTAERGTCFLGEBCCYVYVNSQGIIVTEKVEIRDRIOREABERN 434
DB 447 LAEVVLQNRGLDLTAERGTCFLGEBCCYVYVNSQGIIVTEKVEIRDRIOREABERN 434
QY 435 TGPWGLLSQWMPWILPFLGFLAAIILLLLGPGCIFNLLVNFVSSRIEA 484
DB 507 NPFWTGLHGLPYLLPLRLPLCLLLLTFTGLPIFNKIIAFVKQMDAIQ 556

RESULT 23
Q9GLF7 PRELIMINARY; PRT; 620 AA.

AC Q9GLF7; (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Envelope protein.
GN ENV
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21105294; PubMed=11160757;
RA Baillie G.J., Wilkins R.J.;
RT Possum (Trichosurus vulpecula).
RL J. Virol. 75:2499-2507(2001).
RL EMBL; AF284693; AAG28161.1; -.
DR HSP; P03385; 1MOF.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; E:polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
SQ SEQUENCE 620 AA; 67973 MW; 12E1FF147B805E7F CRC64;

Query Match 20.4%; Score 586; DB 6; Length 620;
Best Local Similarity 29.1%; Pred. No. 2.8e-43;
Matches 156; Conservative 83; Mismatches 173; Indels 124; Gaps 19;

QY 72 CVHSATLCMHANTHY-----W-----TGKMINPSCPG-GLGVTVCW 106

DB 110 CYTSTQCTGKSGTYLTSRQRAYGGSTGDNPTQISGPTNKYACASCDKNIGKNCV 169
QY 107 TYFTQMGSDGGVQDQAREKHVKEVISQLTRV-----HGTSSPY-KGLDLS-KLHET 157
DB 170 SLHAPTHVSDGGGPTDQIREMEVKEVDEIIIRAMYPPLQYHPLALPRGRGVLDVQTYDI 229
QY 158 LRTHRLVSLNLTGLHEVSAQNPTNCWICLPLNFRPVYISIPVPEQNNFSTEINTTS 217
DB 230 LASTHKAINITPDLA-----KDCWLNTL-----GTPMLALLTHDLSFATNC 273
QY 218 VLVGPL-----VSNLEITHSNLTCVKFNTTYTNSQCIKRWTP 257
DB 274 ALSPPFRVQMRPLSAPCIEAFRNSSYDLNVGHASFALCTSNHTFNVTVER-----TP 327
QY 258 PTQIVCLPSGIFVCG-TSAYRLNGSSESMCLFSLVPPMTIY-EQDL-----YSYVIS 311
DB 328 -HLLCPTGPRFVCGNNAFLALPSNWTGCLVQASILEPDINIISGDQVPLPSIDYIAG 385
QY 312 KPRNKEVPILPFGVIGALGTGIGGITTSTQFYKLSOELNGDMRVADSLVTLOQ 371
DB 386 RPK-RAVAFIPLVLGLGVAGMTTGSAGLGVALHSYAKLSNQLINDVQTLSGTIHDLQ 444
QY 372 LNSLAIVLQNRALDLTAERGTCFLGEBCCYVYVNSQGIIVTEKVEIRDRIOREABE 431
DB 445 IDSIAEVVLQNRGLDLTAERGTCFLGEBCCYVYVNSQGIIVTEKVEIRDRIOREABE 431
QY 432 LRNTGFWGLLSQWMPWILPFLGFLAAIILLLLGPGCIFNLLVNFVSSRIEA 482
DB 505 LNNPLNGLHGLPYLLPLRLPLCLLLLTFTGLPIFNKIIAFVKQMDAISARSHIF 564
QY 483 -----VKLQEPKQOS-----KTKLY-----RRPLD---RPASP 508
DB 565 YSRLOQDTTEQOEERLOFTEDLLKEPWHRRWKNLTYALCYKANQPLPLFAEPASP 620

RESULT 24
Q8BI41 PRELIMINARY; PRT; 618 AA.

AC Q8BI41; (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Weakly similar to envelope protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR ENBL; AK086309; BAC39647.1; -.
DR FTR; PT0697; PT0697.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR002050; E:polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
SQ SEQUENCE 618 AA; 69513 MW; 1D4B54588727FC31 CRC64;

Query Match 19.8%; Score 567.5; DB 11; Length 618;
Best Local Similarity 36.7%; Pred. No. 1.2e-41;
Matches 134; Conservative 67; Mismatches 111; Indels 53; Gaps 13;

QY 168 FNTTL-----TGLHEVS-AQNPTNCWICL-----PL-NFRPVYISIPVPEQNNFSTEIN 214
DB 273 YNLTLEPDNSTHPTVMSVNPHTFKVKLQHRDPYFLSHVQPLTGAALSQYSWENEIT 332

```
QY 215 TTSVLGVLAYNLEITHSNTLCVKESNTTYTINSOCIRWVTPPTQIVCLPSGIFVCGT 274
Db 333 -----VQENMDI--TSNI-----FSHLLSFYAFCLN-----SSGVFLCGT 367
QY 275 SAYRCLNGSSSEMCFLSFLVPPMTIYTEQD-----LYSYVISKP---RNKRV-ILPFI 325
Db 368 STVICLPANWGVCTLVFQYDIELLPNNQTVVPVPLFASVLSSDSVLRPKRSPHLPFLA 427
QY 326 GAGVLGALGTGIGITTSQFYKLSOELNGDMERVADSLVTLQDQNSLAAYVLQNR 385
Db 428 GLGSSALGTGAGLATSTLYFOOLSKVLSSETLEEIAASTITLQNSIDSLAGVVLQNR 487
QY 386 LDLLTAERGTCFLGEECCYVYVQSGIVTEKVEIRDRIRORAEELRNTGPMGLLSQW 445
Db 488 LDLLTAERGTCFLGEECCYVYVQSGIVTEKVEIRDRIRORAEELRNTGPMGLLSQW 547
QY 446 PWLPFLGCPAAIILLFPGCFNLLVNFVSSRI-----EAVKLOMEPMQSKTIYR 500
Db 548 FWTTPFGPLFLFLLFTFGCLNCLTRFVSORLGSFVQDTAKRHVDSILQN-FOYKKL 606
QY 501 PLDRP 505
Db 607 PQDSP 611
```

RESULT 25

```
Q9UNN3 PRELIMINARY; PRT; 584 AA.
AC Q9UNN3
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Env protein.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9296739; PubMed=10366582;
RA Lindeskog M., Mager D.L., Blomberg J.;
RT "Isolation of a human endogenous retroviral HERV-H element with an
RT open env reading frame.";
RL Virology 258:441-450(1999).
DR EMBL; AF108843; AAD34324.1; --
DR HSP; P03385; INOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 2.
SQ SEQUENCE 584 AA; 64298 MW; CB41C55A3624FD0A CRC64;
```

Query Match 19.7%; Score 564.5; DB 4; Length 584;

Best Local Similarity 27.1%; Pred.No. 2.1e-41;

Matches 152; Conservative 69; Mismatches 120; Indels 221; Gaps 17;

```
QY 152 SKLHETLR-THRLVSLFNTTGLGHEVSAQNPT---NCWICLPLNFRPVVSPV----- 202
Db 40 SYLHHINTSL------AASNLSLVNVCWLCISLSSAYTAVPALQTDW 85
QY 203 -----PEQ-----WNNFSTEINT 215
Db 86 ATSPISLHRTSFNSPHYPEELIYFLDRSSKTPDISHQAAALLRYLKNLSFYNS 145
QY 216 TSVLGVGLVSNLEITHSNTLCVKESNTT-----YTTNSQ----- 250
Db 146 TPPIGLGLTTQTTPVNAPL-CISWQRPTGIPGNLSPKRCSTLHRSFTTNINETICA 204
QY 251 -----CIR---WVT----- 256
Db 205 FQLHITDKPSINTDKLKNISNYCLGRHLPCISLHPWLSPCSDSPRPSSCLLIPSP 264
```

```
QY 257 -----PPTQI----- 261
Db 265 NNSERLLVDTRRFLIHENRTFPSTQLPHQSPLOPLTAALAGSLGVWVQDTPTSTPSHL 324
QY 262 -----VCLPSGIFVCGTSAYRCLNGSSSEMCFLSFLVPPMTIYT-EQDLYSYVISKPR 314
Db 325 FTLHLQFCLAQGLFLOGSSTYMCLPANWTGCTLVFLTPKIQFANGTEELPVLMTPTQ 384
QY 315 NKRVP-PILPFVIGAGVLG---ALGTGIGITTSQFYKLSOELNGDMERVADSLVTLQD 370
Db 385 QKRVIPLIPLMVGJLSASTVALGTGAGISTVMTFRSLNDSFASITDISQTLSVLQA 444
QY 371 QNSLAAYVLQNRRLDILLTAERGTCFLGEECCYVYVQSGIVTEKVEIRDRIRORAE 430
Db 445 QVDSLAAYVLQNRRLDILLTAERGTCFLGEECCYVYVQSGIVTEKVEIRDRIRORAE 504
QY 431 ELRNTG---PWGLLSQWPMILPFLGCPAAIILLFPGCFNLLVNFVSSRIEAVKLOME 488
Db 505 QASNYAEPFWA-LSNWNWSWLPVIVSPLIPIPLLLFGFCIFRLVSQFIQNIQAITNHSI 563
QY 489 PKMQSKTIYRRPL--DRPAS 507
Db 564 RQFELLTSPQYHPLPQDLPSA 584
```

Search completed: April 19, 2004, 10:35:37

Job time : 51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:55:40 ; Search time 23 seconds
(without alignments)

1207.599 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 538

Sequence: 1 MALPVHFLFTVLLPSFLT.....PPEEISAAQPLLRNSAGSS 538

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCITUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 437 | 81.2 | 538 | 4 | US-09-175-928-4 |
| 2 | 67 | 12.5 | 67 | 1 | US-08-686-878A-49 |
| 3 | 67 | 12.5 | 67 | 1 | US-08-721-489-3 |
| 4 | 45 | 8.4 | 493 | 4 | US-08-979-847B-106 |
| 5 | 13 | 2.4 | 162 | 4 | US-08-979-847B-109 |
| 6 | 8 | 1.5 | 77 | 4 | US-08-979-847B-103 |
| 7 | 8 | 1.5 | 370 | 4 | US-09-252-991A-23512 |
| 8 | 8 | 1.5 | 433 | 3 | US-08-691-563C-87 |
| 9 | 8 | 1.5 | 433 | 4 | US-09-374-766-87 |
| 10 | 8 | 1.5 | 433 | 4 | US-08-979-847B-81 |
| 11 | 7 | 1.3 | 7 | 3 | US-08-836-480-8 |
| 12 | 7 | 1.3 | 7 | 3 | US-09-562-897-8 |
| 13 | 7 | 1.3 | 8 | 3 | US-08-836-480-23 |
| 14 | 7 | 1.3 | 8 | 3 | US-08-836-480-24 |
| 15 | 7 | 1.3 | 8 | 3 | US-09-562-897-23 |
| 16 | 7 | 1.3 | 8 | 3 | US-09-562-897-24 |
| 17 | 7 | 1.3 | 9 | 3 | US-08-836-480-25 |
| 18 | 7 | 1.3 | 9 | 3 | US-08-750-142B-52 |
| 19 | 7 | 1.3 | 9 | 3 | US-09-562-897-25 |
| 20 | 7 | 1.3 | 10 | 3 | US-08-836-480-26 |
| 21 | 7 | 1.3 | 10 | 3 | US-08-750-142B-53 |
| 22 | 7 | 1.3 | 10 | 3 | US-09-562-897-26 |
| 23 | 7 | 1.3 | 12 | 1 | US-07-977-672-1 |
| 24 | 7 | 1.3 | 12 | 4 | US-08-822-940-1 |
| 25 | 7 | 1.3 | 13 | 3 | US-08-836-480-28 |
| 26 | 7 | 1.3 | 13 | 3 | US-08-750-142B-55 |
| 27 | 7 | 1.3 | 13 | 3 | US-09-562-897-28 |

| | | | | | | |
|-----|---|-----|-----|---|----------------------|--------------------|
| 28 | 7 | 1.3 | 15 | 5 | PCT-US93-06751-39 | Sequence 39, Appl |
| 29 | 7 | 1.3 | 20 | 6 | 5204326-31 | Patent No. 5204326 |
| 30 | 7 | 1.3 | 21 | 6 | 5204326-42 | Patent No. 5204326 |
| 31 | 7 | 1.3 | 22 | 6 | 5204326-102 | Patent No. 5204326 |
| 32 | 7 | 1.3 | 22 | 6 | 5204326-91 | Patent No. 5204326 |
| 33 | 7 | 1.3 | 29 | 1 | US-08-443-104-8 | Sequence 8, Appl |
| 34 | 7 | 1.3 | 29 | 1 | US-08-238-130-5 | Sequence 5, Appl |
| 35 | 7 | 1.3 | 29 | 1 | US-08-442-859-8 | Sequence 8, Appl |
| 36 | 7 | 1.3 | 29 | 1 | US-08-398-489-8 | Sequence 8, Appl |
| 37 | 7 | 1.3 | 29 | 2 | PCT-US95-05534-8 | Sequence 8, Appl |
| 38 | 7 | 1.3 | 29 | 5 | 5204326-129 | Patent No. 5204326 |
| 39 | 7 | 1.3 | 31 | 6 | 5204326-115 | Patent No. 5204326 |
| 40 | 7 | 1.3 | 31 | 6 | 5204326-121 | Patent No. 5204326 |
| 41 | 7 | 1.3 | 31 | 6 | 5204326-141 | Patent No. 5204326 |
| 42 | 7 | 1.3 | 49 | 2 | US-08-117-952-792 | Sequence 792, App |
| 43 | 7 | 1.3 | 75 | 3 | US-09-042-012-8 | Sequence 8, Appl |
| 44 | 7 | 1.3 | 75 | 3 | US-09-305-086-2 | Sequence 2, Appl |
| 45 | 7 | 1.3 | 75 | 3 | US-09-457-324-8 | Sequence 8, Appl |
| 46 | 7 | 1.3 | 90 | 4 | US-09-247-155-86 | Sequence 86, Appl |
| 47 | 7 | 1.3 | 90 | 4 | US-09-602-877A-98 | Sequence 98, Appl |
| 48 | 7 | 1.3 | 111 | 1 | US-08-466-886-33 | Sequence 33, Appl |
| 49 | 7 | 1.3 | 111 | 1 | US-08-469-617-33 | Sequence 33, Appl |
| 50 | 7 | 1.3 | 122 | 4 | US-09-198-452A-330 | Sequence 330, App |
| 51 | 7 | 1.3 | 122 | 4 | US-09-770-834-4 | Sequence 4, Appl |
| 52 | 7 | 1.3 | 144 | 2 | US-08-341-843B-27 | Sequence 27, Appl |
| 53 | 7 | 1.3 | 144 | 2 | US-08-427-497B-32 | Sequence 32, Appl |
| 54 | 7 | 1.3 | 145 | 2 | US-08-341-843B-15 | Sequence 15, Appl |
| 55 | 7 | 1.3 | 145 | 2 | US-08-427-497B-20 | Sequence 20, Appl |
| 56 | 7 | 1.3 | 151 | 3 | US-09-188-930-376 | Sequence 276, App |
| 57 | 7 | 1.3 | 151 | 4 | US-09-312-283C-276 | Sequence 276, App |
| 58 | 7 | 1.3 | 152 | 4 | US-09-120-653D-7 | Sequence 7, Appl |
| 59 | 7 | 1.3 | 164 | 4 | US-09-252-991A-16689 | Sequence 16689, A |
| 60 | 7 | 1.3 | 176 | 4 | US-09-134-001C-3181 | Sequence 3181, Ap |
| 61 | 7 | 1.3 | 191 | 4 | US-09-540-2336-2276 | Sequence 2276, Ap |
| 62 | 7 | 1.3 | 192 | 3 | US-08-486-099-107 | Sequence 107, App |
| 63 | 7 | 1.3 | 192 | 3 | US-08-360-107A-117 | Sequence 117, App |
| 64 | 7 | 1.3 | 192 | 3 | US-08-484-223B-107 | Sequence 107, App |
| 65 | 7 | 1.3 | 192 | 3 | US-08-919-597-107 | Sequence 107, App |
| 66 | 7 | 1.3 | 192 | 3 | US-08-475-668A-107 | Sequence 107, App |
| 67 | 7 | 1.3 | 192 | 3 | US-08-485-551A-107 | Sequence 107, App |
| 68 | 7 | 1.3 | 192 | 3 | US-08-471-913A-107 | Sequence 107, App |
| 69 | 7 | 1.3 | 192 | 3 | US-08-485-264A-107 | Sequence 107, App |
| 70 | 7 | 1.3 | 192 | 4 | US-08-474-349A-107 | Sequence 107, App |
| 71 | 7 | 1.3 | 192 | 4 | US-08-470-896-107 | Sequence 107, App |
| 72 | 7 | 1.3 | 192 | 4 | US-08-485-546A-107 | Sequence 107, App |
| 73 | 7 | 1.3 | 196 | 2 | US-08-484-126-7 | Sequence 7, Appl |
| 74 | 7 | 1.3 | 196 | 4 | US-09-374-909-7 | Sequence 7, Appl |
| 75 | 7 | 1.3 | 212 | 2 | US-08-753-159A-4 | Sequence 4, Appl |
| 76 | 7 | 1.3 | 212 | 3 | US-09-133-735-4 | Sequence 4, Appl |
| 77 | 7 | 1.3 | 214 | 4 | US-09-328-352-4246 | Sequence 4246, Ap |
| 78 | 7 | 1.3 | 232 | 4 | US-09-555-352-10 | Sequence 10, Appl |
| 79 | 7 | 1.3 | 268 | 3 | US-08-652-877-4 | Sequence 4, Appl |
| 80 | 7 | 1.3 | 268 | 3 | US-08-476-515A-4 | Sequence 4, Appl |
| 81 | 7 | 1.3 | 269 | 3 | US-08-652-877-5 | Sequence 5, Appl |
| 82 | 7 | 1.3 | 324 | 4 | US-08-476-515A-5 | Sequence 5, Appl |
| 83 | 7 | 1.3 | 329 | 4 | US-09-489-039A-8593 | Sequence 8593, Ap |
| 84 | 7 | 1.3 | 344 | 4 | US-09-489-039A-7441 | Sequence 7441, Ap |
| 85 | 7 | 1.3 | 355 | 4 | US-09-252-991A-23936 | Sequence 23936, A |
| 86 | 7 | 1.3 | 373 | 4 | US-09-543-681A-42933 | Sequence 42933, Ap |
| 87 | 7 | 1.3 | 373 | 4 | US-09-252-991A-30787 | Sequence 30787, A |
| 88 | 7 | 1.3 | 378 | 4 | US-09-540-236-3683 | Sequence 2683, Ap |
| 89 | 7 | 1.3 | 397 | 4 | US-08-506-296B-59 | Sequence 59, Appl |
| 90 | 7 | 1.3 | 397 | 4 | US-09-543-681A-5521 | Sequence 5521, Ap |
| 91 | 7 | 1.3 | 427 | 4 | US-08-506-296B-56 | Sequence 56, Appl |
| 92 | 7 | 1.3 | 437 | 4 | US-09-252-991A-18636 | Sequence 18636, A |
| 93 | 7 | 1.3 | 443 | 4 | US-09-252-991A-21256 | Sequence 21256, A |
| 94 | 7 | 1.3 | 446 | 4 | US-09-543-681A-5864 | Sequence 5864, Ap |
| 95 | 7 | 1.3 | 457 | 4 | US-09-721-870-26 | Sequence 26, Appl |
| 96 | 7 | 1.3 | 466 | 4 | US-09-684-855-113 | Sequence 113, App |
| 97 | 7 | 1.3 | 466 | 4 | US-09-684-855-136 | Sequence 136, App |
| 98 | 7 | 1.3 | 466 | 4 | US-09-684-855-159 | Sequence 159, App |
| 99 | 7 | 1.3 | 472 | 4 | US-09-328-352-6870 | Sequence 6870, Ap |
| 100 | 7 | 1.3 | 477 | 4 | US-09-489-039A-10570 | Sequence 10570, A |

ALIGNMENTS

```
RESULT 1
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.AJ172A
; CURRENT APPLICATION NUMBER: US/09/175.928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-175-928-4

Query Match      81.2%; Score 437; DB 4; Length 538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALPHYHFLFTVLLPSLTITAPPCCRCMTSSPYQEFLLRMQPCGNIDAPSVRLSKGTP 60
DB 1 MALPHYHFLFTVLLPSLTITAPPCCRCMTSSPYQEFLLRMQPCGNIDAPSVRLSKGTP 60

QY 61 TFTAHTMPCNCHSATLCHMANTHYWTGKMINPCSGGLGVTVCTWYFTQTGMSDGGV 120
DB 61 TFTAHTMPCNCHSATLCHMANTHYWTGKMINPCSGGLGVTVCTWYFTQTGMSDGGV 120

QY 121 QDQAREKHVEVISQLTRVHGTSPPYKGLDLKSLKHLTHRLVLSLNTLTGLHEVSA 180
DB 121 QDQAREKHVEVISQLTRVHGTSPPYKGLDLKSLKHLTHRLVLSLNTLTGLHEVSA 180

QY 181 QNPTNCWICLPNRPVYSIPVPOQWNNFSTEINTTSVLGVLVSNLEITHTNLCVKF 240
DB 181 QNPTNCWICLPNRPVYSIPVPOQWNNFSTEINTTSVLGVLVSNLEITHTNLCVKF 240

QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFPVCGTSAYRCLNGSGSEMCFLSFLVPPMTIY 300
DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFPVCGTSAYRCLNGSGSEMCFLSFLVPPMTIY 300

QY 301 TEQDLYSVISKPNKRVPIIPFVIGAGVLGALGTGGITTSQFYKLSQELNGDME 360
DB 301 TEQDLYSVISKPNKRVPIIPFVIGAGVLGALGTGGITTSQFYKLSQELNGDME 360

QY 361 VADSLVTLQDQNSLAADVLRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVE 420
DB 361 VADSLVTLQDQNSLAADVLRALDILLTAERGCTCLFLGEECCYYVNSGIVTEKVE 420

QY 421 IRDRIQRAEELRNTGPNGLLSQWNPWILPFLGPLAAIILLLFGPCIFNLLVNFSSRI 480
DB 421 IRDRIQRAEELRNTGPNGLLSQWNPWILPFLGPLAAIILLLFGPCIFNLLVNFSSRI 480

QY 481 EAVKLQEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS 538
DB 481 EAVKLQEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS 538
```

RESULT 2

```
US-08-686-878A-49
; Sequence 49, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686.878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 67 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-686-878A-49

Query Match      12.5%; Score 67; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 1e-56;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPHYHFLFTVLLPSLTITAPPCCRCMTSSPYQEFLLRMQPCGNIDAPSVRLSKGTP 60
DB 1 MALPHYHFLFTVLLPSLTITAPPCCRCMTSSPYQEFLLRMQPCGNIDAPSVRLSKGTP 60

QY 61 TFTAHTH 67
DB 61 TFTAHTH 67

RESULT 3
US-08-721-489-3
; Sequence 3, Application US/08721489
; Patent No. 5786465
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 71
```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 67 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-721-489-3

Query Match 12.5%; Score 67; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 1e-56;
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALPYHIFLTVLLPFTLTAPPCCRCMTSSSYQBFELRMORPGNIDAPSVRSLSKGTP 60
DB 1 MALPYHIFLTVLLPFTLTAPPCCRCMTSSSYQBFELRMORPGNIDAPSVRSLSKGTP 60
QY 61 TTTAHTH 67
DB 61 TTTAHTH 67

RESULT 4
US-08-979-847B-106
Sequence 106, Application US/08979847B
Patent No. 6582703
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-08-979-847B-106

Query Match 8.4%; Score 45; DB 4; Length 493;
Best Local Similarity 100.0%; Pred. No. 8.6e-35;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 199 SIPVPEQWNFSTEINTSVLGVPLVSNLEITHTSNLTGVKFSNT 243
DB 199 SIPVPEQWNFSTEINTSVLGVPLVSNLEITHTSNLTGVKFSNT 243

RESULT 5
US-08-979-847B-109
Sequence 109, Application US/08979847B
Patent No. 6582703
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-08-979-847B-109

Query Match 2.4%; Score 13; DB 4; Length 162;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 RGGTCLFLGEECC 405
DB 13 RGGTCLFLGEECC 25

RESULT 6
US-08-979-847B-103
; Sequence 103, Application US/08979847B
; Patent No. 6582703

GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
BESEME, FREDERIC
BEDIN, FREDERIC
PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUKER, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-08-979-847B-103

Query Match 1.5%; Score 8; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 PCIFNLLV 473
DB 1 PCIFNLLV 8

RESULT 7

US-09-252-991A-23512
; Sequence 23512, Application US/09252991A
; Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107195.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23512
LENGTH: 370
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23512

Query Match 1.5%; Score 8; DB 4; Length 370;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 GAGVLGAL 333
DB 331 GAGVLGAL 338

RESULT 8

US-08-691-563C-87
; Sequence 87, Application US/08691563C
; Patent No. 6001987
GENERAL INFORMATION:

APPLICANT: Herve PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BEDIN
APPLICANT: Glauca PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996

ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-691-563C-87

Query Match 1.5%; Score 8; DB 3; Length 433;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 FVSSRIEA 482
DB 145 FVSSRIEA 152

RESULT 9
US-09-374-766-87
; Sequence 87, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: peptide
US-09-374-766-87

Query Match 1.5%; Score 8; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 FVSSRIEA 482

DB 145 FVSSRIEA 152

RESULT 10
US-08-979-847B-81
; Sequence 81, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 81:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-08-979-847B-81

Query Match 1.5%; Score 8; DB 4; Length 433;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 FVSSRIEA 482
DB 145 FVSSRIEA 152

RESULT 11
US-08-836-480-8
; Sequence 8, Application US/08836480
; Patent No. 6103697
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; APPLICANT: Erickson, Tomas
; APPLICANT: Karabelas, Kostas
; APPLICANT: Lindvall, Magnus

APPLICANT: Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,480
FILING DATE: 14-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 95\403256-8
FILING DATE: 14-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-8906
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-480-8

Query Match 1.3%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 12
US-08-836-897-8
Sequence 8, Application US/09562897
Patent No. 6228374
GENERAL INFORMATION:
APPLICANT: Bergstrand, Hakan
Erickson, Tomas
Karabellas, Kostas
Lindvall, Magnus
Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/562,897
FILING DATE: 01-May-2000

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,480
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 7 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-562-897-8

Query Match 1.3%; Score 7; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 13
US-08-836-480-23
Sequence 23, Application US/08836480
Patent No. 6103697
GENERAL INFORMATION:
APPLICANT: Bergstrand, Hakan
Erickson, Tomas
Karabellas, Kostas
Lindvall, Magnus
Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,480
FILING DATE: 14-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 95\403256-8
FILING DATE: 14-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-836-480-23

Query Match 1.3%; Score 7; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 14

US-08-836-480-24
; Sequence 24, Application US/08836480
; Patent No. 6103697

GENERAL INFORMATION:

APPLICANT: Bergstrand, Hakan
APPLICANT: Erickson, Tomas
APPLICANT: Karabelas, Kostas
APPLICANT: Lindvall, Magnus
APPLICANT: Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,480
FILING DATE: 14-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 95\403256-8
FILING DATE: 14-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Other
LOCATION: 8...8
OTHER INFORMATION: where Xaa at position 8 is "Leu-NH2"
US-08-836-480-24

Query Match 1.3%; Score 7; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 15

US-09-562-897-23
; Sequence 23, Application US/09562897

; Patent No. 6228374

GENERAL INFORMATION:

APPLICANT: Bergstrand, Hakan
APPLICANT: Erickson, Tomas
APPLICANT: Karabelas, Kostas
APPLICANT: Lindvall, Magnus
APPLICANT: Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/562,897
FILING DATE: 01-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,480
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 8 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 23:
US-09-562-897-23

Query Match 1.3%; Score 7; DB 3; Length 8;

Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

Qy 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 16

US-09-562-897-24
; Sequence 24, Application US/09562897
; Patent No. 6228374

GENERAL INFORMATION:

APPLICANT: Bergstrand, Hakan
APPLICANT: Erickson, Tomas
APPLICANT: Karabelas, Kostas
APPLICANT: Lindvall, Magnus
APPLICANT: Sarnstrand, Bengt
TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804

COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/562.897
  FILING DATE: 01-May-2000
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/836,480
  FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
  NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617-542-5070
  TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 24:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 8 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
  FEATURE:
    NAME/KEY: Other
    LOCATION: 8...8
    OTHER INFORMATION: where Xaa at position 8 is "Leu-NH2"
  SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-562-897-24

Query Match      1.3%; Score 7; DB 3; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      462 LLFGPCI 468
DB      1 LLFGPCI 7

RESULT 17
US-08-836-480-25
; Sequence 25, Application US/08836480
; Patent No. 6103697
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; APPLICANT: Erickson, Tomas
; APPLICANT: Karabelas, Kostas
; APPLICANT: Lindvall, Magnus
; APPLICANT: Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,480
; FILING DATE: 14-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 95\403256-8
; FILING DATE: 14-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.

```

```

REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 06275/036001
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617-542-5070
  TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 25:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 9 amino acids
    TYPE: amino acid
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
  FEATURE:
    NAME/KEY: Other
    LOCATION: 9...9
    OTHER INFORMATION: where Xaa at position 9 is "Asn-NH2"
  SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-08-836-480-25

Query Match      1.3%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      462 LLFGPCI 468
DB      1 LLFGPCI 7

RESULT 18
US-08-750-142B-52
; Sequence 52, Application US/08750142B
; Patent No. 6228373
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; APPLICANT: Eriksson, Tomas
; APPLICANT: Lindvall, Magnus
; APPLICANT: Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,142B
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SE96/00365
; FILING DATE: 22-MAR-1996
; APPLICATION NUMBER: SE9501067-4
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/062001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-750-142B-52

```

Query Match 1.3%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05; 0; Indels 0;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 19

US-09-562-897-25
; Sequence 25, Application US/09562897
; Patent No. 6228374
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; Erickson, Tomas
; Karabelas, Kostas
; Lindvall, Magnus
; Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,897
; FILING DATE: 01-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,480
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/036001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-8906
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Other
; LOCATION: 9...9
; OTHER INFORMATION: where Xaa at position 9 is "Asn-NH2"
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-09-562-897-25
1.3%; Score 7; DB 3; Length 9;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 20

US-08-836-480-26
; Sequence 26, Application US/08836480
; Patent No. 6103657

GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; Erickson, Tomas
; Karabelas, Kostas
; Lindvall, Magnus
; Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,480
; FILING DATE: 14-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 95\403256-8
; FILING DATE: 14-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/036001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-8906
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-836-480-26

Query Match 1.3%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 21

US-08-750-142B-53
; Sequence 53, Application US/08750142B
; Patent No. 6228373
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; Erickson, Tomas
; Karabelas, Kostas
; Lindvall, Magnus
; Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

US-08-750-142B-53
1.3%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

```

; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq For Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,142B
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/SE96/00365
; FILING DATE: 22-MAR-1996
; APPLICATION NUMBER: SE9501067-4
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/062001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-750-142B-53

Query Match 1.3%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 22
US-03-562-897-26
; Sequence 26, Application US/09562897
; Patent No. 6228374
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; ERICKSON, Tomas
; KARABELAS, Kostas
; LINDVALL, Magnus
; SARNSTRAND, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,897
; FILING DATE: 01-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: 08/836,480
; APPLICATION NUMBER:
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/036001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-562-897-26

Query Match 1.3%; Score 7; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 23
US-07-977-672-1
; Sequence 1, Application US/07977672
; Patent No. 5651993
; GENERAL INFORMATION:
; APPLICANT: Edelson, Richard L.
; APPLICANT: Gasparro, Francis P.
; APPLICANT: Tigelaar, Robert E.
; TITLE OF INVENTION: Specific Immune System Modulation
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Ave.
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977,672
; FILING DATE: 19921118
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: Y0060/7001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/720-3500
; TELEFAX: 617/720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: mouse
; US-07-977-672-1

Query Match 1.3%; Score 7; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 381 QNRRLD 387
Db 4 QNRRLD 10

RESULT 24
US-08-822-940-1

```


;; FILING DATE: 24-MAR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fraser, Janis K.
;; REGISTRATION NUMBER: 34,819
;; REFERENCE/DOCKET NUMBER: 06275/062001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617-542-5070
;; TELEFAX: 617-542-8906
;; INFORMATION FOR SEQ ID NO: 55:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 13 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-750-142B-55

Query Match 1.3%; Score 7; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 27
US-09-562-897-28
; Sequence 28, Application US/09562897
; Patent No. 6228374
; GENERAL INFORMATION:
; APPLICANT: Bergstrand, Hakan
; Erickson, Tomas
; Karabelas, Kostas
; Lindvall, Magnus
; Sarnstrand, Bengt
; TITLE OF INVENTION: NEW PEPTIDES WITH IMMUNOMODULATORY EFFECTS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,897
; FILING DATE: 01-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,480
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 06275/036001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Other
; LOCATION: 13...13
; OTHER INFORMATION: where Xaa at position 13 is "Glu-NH2"

;; SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-09-562-897-28

Query Match 1.3%; Score 7; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 LLFGPCI 468
Db 1 LLFGPCI 7

RESULT 28
PCT-US93-06751-39
; Sequence 39, Application PC/TUS9306751
; GENERAL INFORMATION:
; APPLICANT: P. Keller, A. J. Conley, A.R. Shaw, B.A. Arnold
; TITLE OF INVENTION: Immunological Conjugates of OMPc and
; TITLE OF INVENTION: HIV-Specific Selected Principal Neutralization GxG Epitopes
; NUMBER OF SEQUENCES: 146
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06751
; FILING DATE: 19930719
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Meredith, Roy D.
; REGISTRATION NUMBER: 30,777
; REFERENCE/DOCKET NUMBER: 18614
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-4678
; TELEFAX: (908) 594-4720
; TELEX: 138825
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE: Random Epitope Library Alpha
PCT-US93-06751-39

Query Match 1.3%; Score 7; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 TSVLVGP 222
Db 4 TSVLVGP 10

RESULT 29
5204326-31
; Patent No. 5204326
; APPLICANT: FUJII, SETSURO;YAMAMOTO, YOSHITO;SHIMIZU, FUMIO
; INARI, MASATOSHI;KINOSHITA, NAOSUMI;NAKAMURA, SHIZUO;HIROHASHI,
; MITSURU; SAKAMOTO, TAKASHI;TSUTSUMI, KAZUHIKO;SHIRASAKA, TETSUHIKO
; TITLE OF INVENTION: POLYPEPTIDE DERIVATIVES AND CALCIUM
; METABOLISM IMPROVING AGENT

```
; NUMBER OF SEQUENCES: 147
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/493,359
; FILING DATE: 14-MAR-1990
; SEQ ID NO:31:
; LENGTH: 20
5204326-31

Query Match      1.3%; Score 7; DB 6; Length 20;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 KLSQELN 355
Db 1 KLSQELN 7

RESULT 30
5204326-42
; Patent No. 5204326
; APPLICANT: FUJII, SETSURO;YAMAMOTO, YOSHIHIITO;SHIMIZU, FUMIO
; INAI, MASATOSHI;KINOSHITA, NAOSUMI;NAKAMURA, SHIZUO;HIROHASHI,
; MITSURU; SAKAMOTO, TAKASHI;TSUTSUMI, KAZUHIKO;SHIRASAKA, TETSUHIKO
; TITLE OF INVENTION: POLYPEPTIDE DERIVATIVES AND CALCIUM
; METABOLISM IMPROVING AGENT
; NUMBER OF SEQUENCES: 147
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/493,359
; FILING DATE: 14-MAR-1990
; SEQ ID NO:42:
; LENGTH: 21
5204326-42

Query Match      1.3%; Score 7; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 KLSQELN 355
Db 1 KLSQELN 7
```

Search completed: April 19, 2004, 11:01:59
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:59:20 ; Search time 45 Seconds

(without alignments)
3772.195 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 538

Sequence: 1 MALPYHFLFTVLLPSFTLT.....PPEISAAQPLLRNSAGSS 538

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database:

SPREMBL 25:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 538 | 100.0 | 538 | 4 | Q9UQF0 |
| 2 | 437 | 81.2 | 538 | 4 | Q9NKG3 |
| 3 | 437 | 81.2 | 538 | 4 | Q9NRZ2 |
| 4 | 432 | 80.3 | 533 | 4 | Q8NHY7 |
| 5 | 253 | 47.0 | 410 | 4 | Q95244 |
| 6 | 180 | 33.5 | 180 | 4 | Q96TB5 |
| 7 | 170 | 31.6 | 319 | 4 | Q96TB4 |
| 8 | 170 | 31.6 | 319 | 4 | Q96TB4 |
| 9 | 150 | 27.9 | 242 | 4 | Q96L59 |
| 10 | 90 | 16.7 | 319 | 4 | Q95245 |
| 11 | 53 | 9.9 | 74 | 6 | Q8MIQ0 |
| 12 | 52 | 9.7 | 52 | 4 | Q9UQF1 |
| 13 | 47 | 8.7 | 180 | 4 | Q96TB4 |
| 14 | 45 | 8.4 | 542 | 15 | Q91W9 |
| 15 | 38 | 7.1 | 179 | 4 | Q96TB8 |
| 16 | 37 | 6.9 | 180 | 4 | Q96TB1 |

| | | | | | | |
|----|----|-----|------|----|--------|--------------------|
| 17 | 37 | 6.9 | 180 | 4 | Q96TB6 | Q96tb6 homo sapien |
| 18 | 27 | 5.0 | 180 | 4 | Q96TB2 | Q96tb2 homo sapien |
| 19 | 25 | 4.6 | 179 | 4 | Q96TB3 | Q96tb3 homo sapien |
| 20 | 25 | 4.6 | 180 | 4 | Q96TB7 | Q96tb7 homo sapien |
| 21 | 18 | 3.3 | 51 | 6 | Q8MK65 | Q8mk65 pan troglod |
| 22 | 11 | 2.0 | 144 | 4 | Q8NC12 | Q8nc12 homo sapien |
| 23 | 11 | 2.0 | 467 | 4 | Q00354 | Q00354 homo sapien |
| 24 | 11 | 2.0 | 555 | 15 | Q9N2J9 | Q9n2j9 herv-h/env6 |
| 25 | 11 | 2.0 | 584 | 4 | Q9UNM3 | Q9unm3 homo sapien |
| 26 | 11 | 2.0 | 584 | 6 | Q8MIB6 | Q8mib6 pan troglod |
| 27 | 11 | 2.0 | 584 | 15 | Q9N2K0 | Q9n2k0 herv-h/env6 |
| 28 | 11 | 2.0 | 618 | 11 | Q8BI41 | Q8bi41 mus musculu |
| 29 | 10 | 1.9 | 129 | 4 | Q8NHM1 | Q8nhw1 homo sapien |
| 30 | 10 | 1.9 | 247 | 11 | Q9D576 | Q9d576 mus musculu |
| 31 | 8 | 1.5 | 287 | 16 | Q88S72 | Q88s72 lactobacill |
| 32 | 8 | 1.5 | 312 | 16 | Q8Z7M6 | Q8z7m6 salmonella |
| 33 | 8 | 1.5 | 328 | 5 | O16953 | O16953 caenorhabdi |
| 34 | 8 | 1.5 | 337 | 10 | Q8S1B5 | Q8s1b5 oryza sativ |
| 35 | 8 | 1.5 | 349 | 16 | Q98MV5 | Q98mv5 rhizobium l |
| 36 | 8 | 1.5 | 354 | 16 | Q7V953 | Q7v953 prochloroco |
| 37 | 8 | 1.5 | 406 | 16 | Q8D1Z6 | Q8di26 synechococc |
| 38 | 8 | 1.5 | 421 | 5 | Q7Z263 | Q7z263 atemia san |
| 39 | 8 | 1.5 | 497 | 10 | Q8H8A8 | Q8h8a8 taxus chine |
| 40 | 8 | 1.5 | 517 | 4 | Q96N86 | Q96n86 homo sapien |
| 41 | 8 | 1.5 | 521 | 4 | Q8N9M2 | Q8n9m2 homo sapien |
| 42 | 8 | 1.5 | 524 | 3 | Q87208 | Q872q8 neurospora |
| 43 | 8 | 1.5 | 530 | 11 | Q8BK46 | Q8bk46 mus musculu |
| 44 | 8 | 1.5 | 563 | 15 | Q9N2J9 | Q9n2j9 herv-h/env6 |
| 45 | 8 | 1.5 | 580 | 8 | Q8M6U4 | Q8m6u4 tigrionus j |
| 46 | 8 | 1.5 | 696 | 16 | Q88M8 | Q88hm8 pseudomonas |
| 47 | 8 | 1.3 | 1077 | 3 | O74853 | O74853 schizosacch |
| 48 | 7 | 1.3 | 21 | 15 | O05604 | O05604 human endog |
| 49 | 7 | 1.3 | 46 | 16 | Q8FIE0 | Q8fi0 escherichia |
| 50 | 7 | 1.3 | 70 | 6 | Q8QF8 | Q8qf8 eulemur ful |
| 51 | 7 | 1.3 | 71 | 5 | Q8MY05 | Q8my05 caenorhabdi |
| 52 | 7 | 1.3 | 88 | 16 | Q88WD4 | Q88wd4 lactobacill |
| 53 | 7 | 1.3 | 90 | 4 | Q96DR8 | Q96dr8 homo sapien |
| 54 | 7 | 1.3 | 95 | 16 | Q9RTF8 | Q9rtf8 deinococcus |
| 55 | 7 | 1.3 | 95 | 16 | Q87EQ8 | Q87eq8 xylella fas |
| 56 | 7 | 1.3 | 97 | 11 | Q62102 | Q62102 mus musculu |
| 57 | 7 | 1.3 | 101 | 16 | Q9RWH2 | Q9rwh2 deinococcus |
| 58 | 7 | 1.3 | 103 | 15 | Q85615 | Q85615 kirsten mur |
| 59 | 7 | 1.3 | 103 | 16 | Q89CB6 | Q89cb6 bradyrhizob |
| 60 | 7 | 1.3 | 106 | 16 | Q8KFD4 | Q8kfd4 chlorobium |
| 61 | 7 | 1.3 | 108 | 10 | Q94HX3 | Q94hx3 oryza sativ |
| 62 | 7 | 1.3 | 108 | 10 | Q7XCR4 | Q7xcr4 oryza sativ |
| 63 | 7 | 1.3 | 110 | 11 | Q61977 | Q61977 mus musculu |
| 64 | 7 | 1.3 | 112 | 16 | Q9PF11 | Q9pf11 xylella fas |
| 65 | 7 | 1.3 | 115 | 15 | Q7ZEL2 | Q7zel2 human immun |
| 66 | 7 | 1.3 | 122 | 11 | Q8VEA7 | Q8vea7 mus musculu |
| 67 | 7 | 1.3 | 123 | 2 | Q8VMA3 | Q8vma3 xanthomonas |
| 68 | 7 | 1.3 | 124 | 15 | Q9WLW8 | Q9wlw8 friend sple |
| 69 | 7 | 1.3 | 127 | 10 | O22942 | O22942 arabidopsis |
| 70 | 7 | 1.3 | 130 | 15 | Q77246 | Q77246 human immun |
| 71 | 7 | 1.3 | 130 | 15 | Q77239 | Q77239 human immun |
| 72 | 7 | 1.3 | 130 | 15 | Q77241 | Q77241 human immun |
| 73 | 7 | 1.3 | 130 | 15 | Q77247 | Q77247 human immun |
| 74 | 7 | 1.3 | 130 | 15 | Q77242 | Q77242 human immun |
| 75 | 7 | 1.3 | 130 | 15 | Q77240 | Q77240 human immun |
| 76 | 7 | 1.3 | 130 | 15 | Q77243 | Q77243 human immun |
| 77 | 7 | 1.3 | 130 | 15 | Q77245 | Q77245 human immun |
| 78 | 7 | 1.3 | 132 | 2 | Q7X322 | Q7x322 pseudomonas |
| 79 | 7 | 1.3 | 133 | 17 | Q97VC6 | Q97vc6 sulfolobus |
| 80 | 7 | 1.3 | 140 | 2 | Q8VQY9 | Q8vqy9 myxococcus |
| 81 | 7 | 1.3 | 140 | 5 | Q9XWY9 | Q9xw9 caenorhabdi |
| 82 | 7 | 1.3 | 141 | 16 | Q7U876 | Q7u876 synechococc |
| 83 | 7 | 1.3 | 144 | 16 | Q9KPI9 | Q9kpi9 vibrio chol |
| 84 | 7 | 1.3 | 147 | 2 | O85781 | O85781 myxococcus |
| 85 | 7 | 1.3 | 171 | 8 | O20597 | O20597 calyptogena |
| 86 | 7 | 1.3 | 171 | 8 | O20578 | O20578 calyptogena |
| 87 | 7 | 1.3 | 171 | 11 | Q8R504 | Q8r504 mus musculu |
| 88 | 7 | 1.3 | 174 | 16 | Q8CSG9 | Q8csg9 staphylococ |
| 89 | 7 | 1.3 | 181 | 16 | Q92ZC2 | Q92zc2 rhizobium m |

90 Q85641 moloney mur
91 Q64107 rattus sp.
92 Q96786 cooperia on
93 Q85640 murine leuk
94 Q83401 moloney mur
95 Q96h29 homo sapien
96 Q88303 simian srv-
97 Q91tk0 arabidopsis
98 Q96d42 homo sapien
99 Q23393 caenorhabdi
100 Q02754 saccharomyc

ALIGNMENTS

RESULT 1

Q9UQF0 PRELIMINARY; PRT; 538 AA.
ID Q9UQF0
AC Q9UQF0
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Envelope protein precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99099005; PubMed=9882319;
RA Blond J.L., Besene F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.;
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072506; AAD14546.2; -
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 538 AA; 59866 MW; C54648A3C7043870 CRC64;

Query Match 100.0%; Score 538; DB 4; Length 538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MALPHYHIFLTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
DB 1 MALPHYHIFLTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
QY 61 TFTAHTMPCNCHSATLCEANHTHYWTKMKNPSCPGGLGVTCWYFTQTMGSDGGV 120
DB 61 TFTAHTMPCNCHSATLCEANHTHYWTKMKNPSCPGGLGVTCWYFTQTMGSDGGV 120
QY 121 QDAREKHKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRVLSLNTLTGLHEVSA 180
DB 121 QDAREKHKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRVLSLNTLTGLHEVSA 180
QY 181 QNFTNCWICLPLNFRPYVSIPEVQNNFSTEINTSVLGVPLVSNLEITHTSNLTCKVF 240
DB 181 QNFTNCWICLPLNFRPYVSIPEVQNNFSTEINTSVLGVPLVSNLEITHTSNLTCKVF 240
QY 241 SNTTWTNSOCHRWTPPTQIVCLPSGIFFFVCGTSAYFCLNGSSSFCFLSFLVPPNTIY 300
DB 241 SNTTWTNSOCHRWTPPTQIVCLPSGIFFFVCGTSAYFCLNGSSSFCFLSFLVPPNTIY 300
QY 301 TEQDLYSVISKPRNKRVPILPFVIGAGVLTGIGITTTSTQFYFKLSOELNGDWER 360

RESULT 2

Q9NZG3 PRELIMINARY; PRT; 538 AA.
ID Q9NZG3
AC Q9NZG3
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Syncytin precursor (Enverin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20155476; PubMed=10693809;
RA Sha M., Lee X., Li X., Veldman G.M., Finnerty H., Racie L.,
RA LaVallie E., Tang X., Edouard P., Howes S., Keith J.C. Jr.,
RA McCoy J.M.;
RT "Syncytin is a captive retroviral envelope protein involved in human
RT placental morphogenesis.";
RL Nature 403:785-789(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
RA Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
RT its env transcripts.";
RL C. R. Acad. Sci. III, Sci. Vie 312:857-863(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=21995840; PubMed=11990458;
RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7q family in the human genome. Potential for protein
RT expression and gene regulation.";
RL Cell. Mol. Biol. 48:213-217(2002).
DR EMBL; AF208161; AAF28334.1; -
DR EMBL; AF513360; AAM47599.1; -
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 538 AA; 59893 MW; B9498850CB34FB78 CRC64;

Query Match 81.2%; Score 437; DB 4; Length 538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALPHYHIFLTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
DB 1 MALPHYHIFLTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60

QY 61 TFTAHTMPNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWYFTQTGMSDGGV 120
Db 61 TFTAHTMPNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWYFTQTGMSDGGV 120
QY 121 QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLEHETLRTHTRLVSLFNLTGLHEVSA 180
Db 121 QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLEHETLRTHTRLVSLFNLTGLHEVSA 180
QY 181 QNPTNCWICLPLNFRPVSIPVPEQWNNFTEINTSVLVGPLVSNLEITHSTNLTVCVF 240
Db 181 QNPTNCWICLPLNFRPVSIPVPEQWNNFTEINTSVLVGPLVSNLEITHSTNLTVCVF 240
QY 241 SNTTYTNSQICRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
Db 241 SNTTYTNSQICRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
QY 301 TEQDLYSVVSKPRNKRPVILPFVIGAGVGLGTTGGITTTSTQFYKLSQELNGDME 360
Db 301 TEQDLYSVVSKPRNKRPVILPFVIGAGVGLGTTGGITTTSTQFYKLSQELNGDME 360
QY 361 VADSLVTLQDLSLAAVLQNRALDLLTAERGTCFLGEECCYVYVQSGIVTEKVE 420
Db 361 VADSLVTLQDLSLAAVLQNRALDLLTAERGTCFLGEECCYVYVQSGIVTEKVE 420
QY 421 IRDRIQRAEELNRTGPMGLLSQMPWILPFLGPLAAIILLFPGCIENLLVNFVSSRI 480
Db 421 IRDRIQRAEELNRTGPMGLLSQMPWILPFLGPLAAIILLFPGCIENLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVDNDIKGTPPEEISAAQPLRPNSAGSS 538
Db 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVDNDIKGTPPEEISAAQPLRPNSAGSS 538

RESULT 3
Q9NRZ2 PRELIMINARY; PRT; 538 AA.
AC Q9NRZ2
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20284713; PubMed=10826480;
RA Volset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
Paranhos-Baccala G.,
RT "Chromosomal distribution and coding capacity of the human endogenous
retrovirus HERV-W family."
RL AIDS Res. Hum. Retroviruses 16:731-740(2000).
DR EMBL; AF156963; AAF74215.1; -
DR HSSP; P03385; 1MOF.
DR Genew; HGNC:13525; ERVWEL.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
SQ SEQUENCE 538 AA; 59894 MW; 476998A13A2A256F CRC64;

Query Match 81.2%; Score 437; DB 4; Length 538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPQEFELWRCQGNIDAPSYSLSGTIP 60
Db 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPQEFELWRCQGNIDAPSYSLSGTIP 60
QY 61 TFTAHTMPNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWYFTQTGMSDGGV 120

Db 61 TFTAHTMPNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWYFTQTGMSDGGV 120
QY 121 QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLEHETLRTHTRLVSLFNLTGLHEVSA 180
Db 121 QDQAREKHVKVISQLTRVHGTSPPYKGLDLSKLEHETLRTHTRLVSLFNLTGLHEVSA 180
QY 181 QNPTNCWICLPLNFRPVSIPVPEQWNNFTEINTSVLVGPLVSNLEITHSTNLTVCVF 240
Db 181 QNPTNCWICLPLNFRPVSIPVPEQWNNFTEINTSVLVGPLVSNLEITHSTNLTVCVF 240
QY 241 SNTTYTNSQICRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
Db 241 SNTTYTNSQICRWVTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
QY 301 TEQDLYSVVSKPRNKRPVILPFVIGAGVGLGTTGGITTTSTQFYKLSQELNGDME 360
Db 301 TEQDLYSVVSKPRNKRPVILPFVIGAGVGLGTTGGITTTSTQFYKLSQELNGDME 360
QY 361 VADSLVTLQDLSLAAVLQNRALDLLTAERGTCFLGEECCYVYVQSGIVTEKVE 420
Db 361 VADSLVTLQDLSLAAVLQNRALDLLTAERGTCFLGEECCYVYVQSGIVTEKVE 420
QY 421 IRDRIQRAEELNRTGPMGLLSQMPWILPFLGPLAAIILLFPGCIENLLVNFVSSRI 480
Db 421 IRDRIQRAEELNRTGPMGLLSQMPWILPFLGPLAAIILLFPGCIENLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVDNDIKGTPPEEISAAQPLRPNSAGSS 538
Db 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVDNDIKGTPPEEISAAQPLRPNSAGSS 538

RESULT 4
Q8NH7 PRELIMINARY; PRT; 533 AA.
AC Q8NH7
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Enverin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99052087; PubMed=9835022;
RA Alliel P.M., Ferin J.P., Pierig R., Nussbaum J.L., Menard A.,
Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
its env transcripts."
RL C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21985840; PubMed=11990458;
RA Alliel P.M., Ferin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7q family in the human genome. Potential for protein
expression and gene regulation."
RL Cell. Mol. Biol. 48:213-217(2002).
DR EMBL; AF506835; RAM33413.1; -
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 533
SQ SEQUENCE 533 AA; 59503 MW; 14FE789BCE1C41CD CRC64;

Query Match 80.3%; Score 432; DB 4; Length 533;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 532; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPQEFELWRCQGNIDAPSYSLSGTIP 60
Db 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPQEFELWRCQGNIDAPSYSLSGTIP 60
QY 61 TFTAHTMPNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCWYFTQTGMSDGGV 120

QY 1 MALPYHIEFTVLLPSFTLTAPPBCRWMTSSPYQEFELWRMORGNIDAPSYRLSKGTP 60
 DB 1 MALPYHIEFTVLLPSFTLTAPPBCRWMTSSPYQEFELWRMORGNIDAPSYRLSKGTP 60
 QY 61 TFTAHTHPRNCYHSATLCHMANTHYWTGKMINFSCPGGLGVTCWYFTQTGMSDGGV 120
 DB 61 TFTAHTHPRNCYHSATLCHMANTHYWTGKMINFSCPGGLGVTCWYFTQTGMSDGGV 120
 QY 121 QDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRLVSLNLTGLHEVSA 180
 DB 121 QDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRLVSLNLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPYVSIIPVEQWNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 240
 DB 181 QNPTNCWICLPLNFRPYVSIIPVEQWNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 240
 QY 241 SNTTNTTNSQCIIRWTPPTQIVCLPSGIFVCGTSAYRCCLNGSSSMCFSLFVPPMTIY 300
 DB 241 SNTTNTTNSQCIIRWTPPTQIVCLPSGIFVCGTSAYRCCLNGSSSMCFSLFVPPMTIY 300
 QY 301 TEQDLYSVYISKPRNKRPVILPFVIGAGVGLGALGTGIGITTSQFYKLSQELNGDME 360
 DB 301 TEQDLYSVYISKPRNKRPVILPFVIGAGVGLGALGTGIGITTSQFYKLSQELNGDME 360
 QY 361 VADSLVTLQDQNSLAAYVQLQNRALDILLTAERGTCFLGEECCYYNQSGIVTEKYKE 420
 DB 361 VADSLVTLQDQNSLAAYVQLQNRALDILLTAERGTCFLGEECCYYNQSGIVTEKYKE 420
 QY 421 IDRIQRAEELNRTGPGWLLSQWNPWILPFLGPLAAIILLFPGFCIFNLLVNFVSSRI 480
 DB 421 IDRIQRAEELNRTGPGWLLSQWNPWILPFLGPLAAIILLFPGFCIFNLLVNFVSSRI 480
 QY 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPBEISAQPLLRPN 533
 DB 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPBEISAQPLLRPN 533

RESULT 5
 Q95244 PRELIMINARY; PRT; 410 AA.
 ID C95244
 AC C95244;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=99099005; PubMed=9882319;
 RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
 RA Mandrand B., Mallet F.;
 RT "Molecular characterization and placental expression of HERV-W, a new
 RT human endogenous retrovirus family";
 RL J. Virol. 73:1175-1185 (1999).
 DR ENBL; AF072505; RAD14545.1;
 DR HSP; P03385; IMOF.
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
 DR GO; GO:0006949; P:syncytium formation; TAS.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_polyprotein; 1.
 FT NON TER 410
 SQ SEQUENCE 410 AA; 45442 MW; 637F1ECE35DE1587 CRC64;
 Query Match 47.0%; Score 253; DB 4; Length 410;
 Best Local Similarity 99.7%; Pred. No. 8.4e-250;
 Matches 353; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 57 KGTPPTAHTHPRNCYHSATLCHMANTHYWTGKMINFSCPGGLGVTCWYFTQTGMSD 116

DB 57 KGTPPTAHTHPRNCYHSATLCHMANTHYWTGKMINFSCPGGLGVTCWYFTQTGMSD 116
 QY 117 GGVQDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRLVSLNLTGLH 176
 DB 117 GGVQDOAREKHVKEVISOLTRVHGTSPPYKGLDLSKLHETLRTLRLVSLNLTGLH 176
 QY 177 EYSAQNPTNCWICLPLNFRPYVSIIPVEQWNNFSTEINTSVLVGLVSNLEITHTSNLT 236
 DB 177 EYSAQNPTNCWICLPLNFRPYVSIIPVEQWNNFSTEINTSVLVGLVSNLEITHTSNLT 236
 QY 237 CVKFSNTTNTTNSQCIIRWTPPTQIVCLPSGIFVCGTSAYRCCLNGSSSMCFSLFVPP 296
 DB 237 CVKFSNTTNTTNSQCIIRWTPPTQIVCLPSGIFVCGTSAYRCCLNGSSSMCFSLFVPP 296
 QY 297 MTITYEQDLYSVYISKPRNKRPVILPFVIGAGVGLGALGTGIGITTSQFYKLSQELNG 356
 DB 297 MTITYEQDLYSVYISKPRNKRPVILPFVIGAGVGLGALGTGIGITTSQFYKLSQELNG 356
 QY 357 DMERVADSLVTLQDQNSLAAYVQLQNRALDILLTAERGTCFLGEECCYYNQ 410
 DB 357 DMERVADSLVTLQDQNSLAAYVQLQNRALDILLTAERGTCFLGEECCYYNQ 410

RESULT 6
 Q96TBS PRELIMINARY; PRT; 180 AA.
 ID Q96TBS
 AC Q96TBS;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21268768; PubMed=11375061;
 RA Kim H.-S., Lee W.-H.;
 RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
 RT identification, and phylogeny";
 RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
 DR ENBL; AB051004; BAB47558.1;
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_Polyprotein; 1.
 FT NON TER 180
 SQ SEQUENCE 180 AA; 19648 MW; 8D2F33813F16090B CRC64;
 Query Match 33.5%; Score 180; DB 4; Length 180;
 Best Local Similarity 100.0%; Pred. No. 2.1e-175;
 Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 NNFSTEINTSVLVGLVSNLEITHTSNLTCKVFSTNTTNSQCIIRWTPPTQIVCLPS 266
 DB 1 NNFSTEINTSVLVGLVSNLEITHTSNLTCKVFSTNTTNSQCIIRWTPPTQIVCLPS 60
 QY 267 GIFVCGTSAYRCCLNGSSSMCFSLFVPPMTIYTEQDLYSVYISKPRNKRPVILPFVIG 326
 DB 61 GIFVCGTSAYRCCLNGSSSMCFSLFVPPMTIYTEQDLYSVYISKPRNKRPVILPFVIG 120
 QY 327 AGVLGALGTGIGITTSQFYKLSQELNGDMERVADSLVTLQDQNSLAAYVQLQNRAL 386
 DB 121 AGVLGALGTGIGITTSQFYKLSQELNGDMERVADSLVTLQDQNSLAAYVQLQNRAL 180

RESULT 7
 Q96L61 PRELIMINARY; PRT; 319 AA.
 ID Q96L61
 AC Q96L61;

```
01-DEC-2001 (TReMBLrel. 19, Created)
01-DEC-2001 (TReMBLrel. 19, Last sequence update)
01-DEC-2001 (TReMBLrel. 19, Last annotation update)
Envelope protein (Fragment).
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
SEQUENCE FROM N.A.
Jern P., Karlsson D., Blomberg J.;
"New HERV-W SU sequences.";
Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AY050299; AAL11493.1; -.
NON_TER 319
SEQUENCE 319 AA; 35655 MW; 7D3397CCCAAD8999 CRC64;

Query Match 31.6%; Score 170; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 6e-165;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
DB 1 MALPHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
QY 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTYFTQTGMSDGGGV 120
DB 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTYFTQTGMSDGGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHLVSLFNT 170
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHLVSLFNT 170

RESULT 8
Q96L59 PRELIMINARY; PRT; 319 AA.
AC Q96L59;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jern P., Karlsson D., Blomberg J.;
RT "New HERV-W SU sequences.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY050301; AAL11495.1; -.
FT NON_TER 319
SQ SEQUENCE 319 AA; 35784 MW; 7D288CCCD1B69999 CRC64;

Query Match 31.6%; Score 170; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 6e-165;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
DB 1 MALPHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPGNIDAPSYRLSKGTP 60
QY 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTYFTQTGMSDGGGV 120
DB 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTYFTQTGMSDGGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHLVSLFNT 170
DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHLVSLFNT 170

RESULT 9
Q96L60 PRELIMINARY; PRT; 242 AA.
AC Q96L60;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Jern P., Karlsson D., Blomberg J.;
RT "New HERV-W SU sequences.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY050300; AAL11494.1; -.
FT NON_TER 319
SQ SEQUENCE 319 AA; 35763 MW; 67822669D1B6899D CRC64;

Query Match 16.7%; Score 90; DB 4; Length 319;
Best Local Similarity 100.0%; Pred. No. 4.6e-83;
Matches 90; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 HANTHYWTGKINPSCPGGLGVTVCTYFTQTGMSDGGGVQDQAREKHVKEVISQLTRVH 140
```

```

Db 81 HANTHWTKMINPSCPGGLGTVTCWYFTQTGMSDGGVQDQAREKHVKVLSQLTRVH 140
QY 141 GTSPGPKGLDLSKHLHETLRTRVLSFNT 170
Db 141 GTSPGPKGLDLSKHLHETLRTRVLSFNT 170

RESULT 11
Q8MIQO PRELIMINARY; PRT; 74 AA.
AC Q8MIQO;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99052087; PubMed=9835022;
RA Alliel P.M., Ferrin J.P., Pierig R., Nussbaum J.L., Menard A.,
RA Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. II. HERV-7g.";
RN C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21985939; PubMed=11990457;
RA Alliel P.M., Ferrin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7g family in the human genome. Potential for protein
RT expression and gene regulation.";
RL Cell. Mol. Biol. 48:205-212(2002).
DR EMBL; AY102609; AM51553.1; -
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR02050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 74 74
FT NON_TER 1
SQ SEQUENCE 74 AA; 7909 MW; 58D9A79BA322E0A4 CRC64;

Query Match 9.9%; Score 53; DB 6; Length 74;
Best Local Similarity 100.0%; Pred. No. 9.4e-46;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 LGTGIGGTTTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRA 385
Db 22 LGTGIGGTTTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRA 74

RESULT 12
Q9UQF1 PRELIMINARY; PRT; 52 AA.
AC Q9UQF1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99099005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.;
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072507; AAD14547.1; -

```

```

FT NON_TER 1
SQ SEQUENCE 52 AA; 5719 MW; 2EDD668523DCB4F4 CRC64;

Query Match 9.7%; Score 52; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 7.3e-45;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 MEPMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 538
Db 1 MEPMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 52

RESULT 13
Q96TB4 PRELIMINARY; PRT; 180 AA.
AC Q96TB4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
OS ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
DR EMBL; AB051507; BAB47559.1; -
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR02050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 180 180
FT NON_TER 180
SQ SEQUENCE 180 AA; 20048 MW; F2B7B2FAF7CE40C3 CRC64;

Query Match 8.7%; Score 47; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 2.9e-39;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 340 ITTSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRA 386
Db 134 ITTSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVVLQNRRA 180

RESULT 14
Q991W9 PRELIMINARY; PRT; 542 AA.
AC Q991W9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Recombinant envelope protein.
OS ENV.
OS Multiple sclerosis associated retrovirus element.
OC Viruses; Retroid viruses; Retroviridae.
OX NCBI_TaxID=89382;
RN [1]
RP SEQUENCE FROM N.A.
RA Perron H., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
RA Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
RA Gebuhrer L.;
RT "Multiple sclerosis retrovirus-induced polyclonal Vb16 T-cell receptor
RT activation as a trigger of abnormal immune response.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF331500; AAK18189.1; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.

```


DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Envelope protein.
SQ SEQUENCE 542 AA; 60625 MW; AAF3A8322E03C9EE CRC64;

Query Match 8.4%; Score 45; DB 15; Length 542;
Best Local Similarity 100.0%; Pred. No. 8.6e-37; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 0;

QY 199 SIPVPQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 243
DB 199 SIPVPQWNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 243

RESULT 15
Q96TB8 PRELIMINARY; PRT; 179 AA.
ID Q96TB8
AC Q96TB8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
DR EMBL; AB050996; BAB47555.1;
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT SEQUENCE 179 AA; 19228 MW; 3A9FACBAG6376E74 CRC64;

Query Match 7.1%; Score 38; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 4.7e-30; Indels 0; Gaps 0;
Matches 38; Conservative 0; Mismatches 0;

QY 266 SGIFFFVCGTSAYKCLNGSSSFCFLSFLVPPMTIYTEQ 303
DB 60 SGIFFFVCGTSAYKCLNGSSSFCFLSFLVPPMTIYTEQ 97

RESULT 16
Q96TB1 PRELIMINARY; PRT; 180 AA.
ID Q96TB1
AC Q96TB1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
DR EMBL; AB050996; BAB47555.1;
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT SEQUENCE 179 AA; 19228 MW; 3A9FACBAG6376E74 CRC64;

Query Match 6.9%; Score 37; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 5e-29; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0;

QY 207 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 243
DB 1 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 37

RESULT 17
Q96TB6 PRELIMINARY; PRT; 180 AA.
ID Q96TB6
AC Q96TB6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization, identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648(2001).
DR EMBL; AB051000; BAB47557.1;
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT SEQUENCE 180 AA; 19656 MW; BF8DFCDB4CAEAF3D CRC64;

Query Match 6.9%; Score 37; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 5e-29; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0;

QY 207 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 243
DB 1 NNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKSNT 37

RESULT 18
Q96TB2 PRELIMINARY; PRT; 180 AA.
ID Q96TB2
AC Q96TB2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;

```

RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB051009; BAB47561.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 19709 MW; 4C8F196C9C788497 CRC64;

Query Match 5.0%; Score 27; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 8.5e-19; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0;

QY 333 LGTGIGGTTSTQFYFKLSQELNGDME 359
DB 127 LGTGIGGTTSTQFYFKLSQELNGDME 153

RESULT 19
Q96TB3 PRELIMINARY; PRT; 179 AA.
AC Q96TB3;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB051008; BAB47560.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT NON_TER 179
SQ SEQUENCE 179 AA; 19471 MW; 70DF3E5658B8E667 CRC64;

Query Match 4.6%; Score 25; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 9.4e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 335 TGIGGTTSTQFYFKLSQELNGDME 359
DB 128 TGIGGTTSTQFYFKLSQELNGDME 152

RESULT 20
Q96TB7 PRELIMINARY; PRT; 180 AA.
AC Q96TB7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21268768; PubMed=11375061;
RA Kim H.-S., Lee W.-H.;
RT "Human endogenous retrovirus HERV-W family: chromosomal localization,
RT identification, and phylogeny.";
RL AIDS Res. Hum. Retroviruses 17:643-648 (2001).
DR EMBL; AB050999; BAB47556.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 19984 MW; 3414BCDD74AC60BA CRC64;

Query Match 4.6%; Score 25; DB 4; Length 180;
Best Local Similarity 100.0%; Pred. No. 9.5e-17;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 LVGPLVSNLEITHSTNLTKVKSNT 243
DB 13 LVGPLVSNLEITHSTNLTKVKSNT 37

RESULT 21
Q8MK65 PRELIMINARY; PRT; 51 AA.
AC Q8MK65;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Envelope glycoprotein (fragment).
GN ENV.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99052087; PubMed=9835022;
RA Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
RA Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. II. HERV-7q.";
RL C. R. Acad. Sci. III, Sci. Vie 321:857-863 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21985839; PubMed=11990457;
RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7q family in the human genome. Potential for protein
RT expression and gene regulation.";
RL Cell. Mol. Biol. 48:205-212 (2002).
DR EMBL; AY098485; AAM28225.1; -.
FT NON_TER 1
FT NON_TER 51
SQ SEQUENCE 51 AA; 5606 MW; 08B5AF8AF612A023 CRC64;

Query Match 3.3%; Score 18; DB 6; Length 51;
Best Local Similarity 100.0%; Pred. No. 4.5e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CLNGSSSMCFSLVPP 296
DB 14 CLNGSSSMCFSLVPP 31

RESULT 22
Q8NC12 PRELIMINARY; PRT; 144 AA.
AC Q8NC12;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)

```

DE Hypothetical protein FLJ90611.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
EN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA ISOGAI T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Sasaki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK075092; BAC11396.1; -
DR GO; GO:0019028; F:structural molecule activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Hypothetical protein.
SQ SEQUENCE 144 AA; 16165 MW; 058C2EE7BF1905A4 CRC64;
Query Match 2.0%; Score 11; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 0.017;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 SLAAVVLQNR 384
DB 14 SLAAVVLQNR 24
RESULT 23
C00354 PRELIMINARY; PRT; 467 AA.
AC O00354;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
EN [1]
RP SEQUENCE FROM N.A.
RA Lindeskog M., Blomberg J.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88902; AAC79121.1; -
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 2.
KW Envelope protein.
FT NON TER 467
SQ SEQUENCE 467 AA; 50913 MW; 4C5C65DA147FBCB4 CRC64;
Query Match 2.0%; Score 11; DB 4; Length 467;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 SLAAVVLQNR 384
DB 448 SLAAVVLQNR 458
RESULT 24
Q9N2J8 PRELIMINARY; PRT; 555 AA.
ID Q9N2J8
AC Q9N2J8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein.
GN ENV.
OS HERV-H/env62
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=129525;
EN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21109114; PubMed=11162811;
RX de Parseval N., Casella J.F., Gressin L., Heidmann T.;
RT "Characterization of the three envelopes of the HERV-H family
encompassing the ISU domain and evolutionary history in primates.";
RL Virology 279:558-569(2001).
DR EMBL; AJ289711; CAB94194.1; -
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
SQ SEQUENCE 555 AA; 60910 MW; E59604545407844 CRC64;
Query Match 2.0%; Score 11; DB 15; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 SLAAVVLQNR 384
DB 448 SLAAVVLQNR 458
RESULT 25
Q9UNM3 PRELIMINARY; PRT; 584 AA.
AC Q9UNM3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Env protein.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
EN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9296739; PubMed=10366582;
RA Lindeskog M., Mager D.L., Blomberg J.;
RT "Isolation of a human endogenous retroviral HERV-H element with an
open env reading frame.";
RL Virology 258:441-450(1999).
DR EMBL; AF108843; AAD34324.1; -
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 2.
SQ SEQUENCE 584 AA; 64238 MW; CB41C55A3624FD0A CRC64;
Query Match 2.0%; Score 11; DB 4; Length 584;
Best Local Similarity 100.0%; Pred. No. 0.058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 374 SLAAVVLQNR 384
DB 448 SLAAVVLQNR 458
RESULT 26
Q8MIB6 PRELIMINARY; PRT; 584 AA.
ID Q8MIB6
AC Q8MIB6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)

| | |
|-----------|--|
| DT | 01-JUN-2003 (TrenBMLrel. 24, Last annotation update) |
| DE | Weakly similar to envelope protein. |
| OS | Mus musculus (Mouse). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| CC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| NBI | _TaxID=10090; |
| [1] | |
| RN | SEQUENCE FROM N.A. |
| RP | STRAIN=C57BL/6J; TISSUE=Head; |
| RC | MEDLINE=22354683; PubMed=12466851; |
| RX | The FANTOM Consortium, |
| RA | the RIKEN Genome Exploration Research Group Phase I & II Team; |
| RT | "Analysis of the mouse transcriptome based on functional annotation of |
| RT | 60,770 full-length cDNAs." |
| RL | Nature 420:563-573(2002)." |
| DR | EMBL; AK098309; BAC39647.1; -. |
| DR | PIR; PT0697; FT0697. |
| DR | GO; GO:0019028; C:viral capsid activity; IEA. |
| DR | GO; GO:0005198; F:structural molecule activity; IEA. |
| DR | InterPro; IP002050; Env_polyprotein. |
| DR | Pfam; PF00429; ENV_polyprotein_1. |
| DR | SEQUENCE 618 AA; 69513 MW; 1D4B5458872FC31 CRC64; |
| SQ | |
| Qy | Query Match 2.0%; Score 11; DB 11; Length 618; |
| Db | Best Local Similarity 100.0%; Pred.No. 0.061; |
| | Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| | |
| Qy | 378 VVLQNRALDL 388 |
| Db | 480 VVLQNRALDL 490 |
| | |
| RESULT 29 | |
| Q8NHW1 | PRELIMINARY; PRT; 129 AA. |
| ID | Q8NHW1 |
| AC | 01-OCT-2002 (TrenBMLrel. 22, Created) |
| DT | 01-OCT-2002 (TrenBMLrel. 22, Last sequence update) |
| DT | 01-OCT-2002 (TrenBMLrel. 22, Last annotation update) |
| DE | Enverin-2 (Fragment). |
| DE | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |
| NBI | _TaxID=9606; |
| [1] | |
| RN | SEQUENCE FROM N.A. |
| RP | TISSUE=Placenta; |
| RC | MEDLINE=99052087; PubMed=9835022; |
| RX | Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A., |
| RA | Rieger F.; |
| RT | "Endogenous retroviruses and multiple sclerosis. II. HERV-7q;" |
| RL | C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998). |
| [2] | |
| RN | SEQUENCE FROM N.A. |
| RP | TISSUE=Placenta; |
| RC | MEDLINE=21985939; PubMed=11990457; |
| RA | Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieser F.; |
| RT | "The HERV-w/q family in the human genome. Potential for protein |
| RT | expression and gene regulation."; |
| RL | Cell. Mol. Biol. 48:205-212(2002). |
| DR | ENBL; AY098482; AAM33141.1; -. |
| FT | NON TER 1 |
| FT | NON TER 129 |
| SQ | SEQUENCE 129 AA; 14217 MW; FP319F083388E84 CRC64; |
| | |
| Qy | Query Match 1.9%; Score 10; DB 4; Length 129; |
| Db | Best Local Similarity 100.0%; Pred.No. 0.16; |
| | Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| | |
| Qy | 477 SSRIEAVKLQ 486 |
| Db | 69 SSRIEAVKLQ 78 |
| | |

| | |
|--------------------|---|
| DT | 01-OCT-2002 (TReMBLrel. 22, Last sequence update) |
| DE | 01-JUN-2003 (TReMBLrel. 24, Last annotation update) |
| DD | Envelope protein. |
| DN | ENV. |
| OS | Pan troglodytes (Chimpanzee). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| CC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan. |
| NCSI_TaxID=9598; | |
| [1] | |
| RN | SEQUENCE FROM N.A. |
| RP | TISSUE=Peripheral blood leukocytes; |
| RC | Besit L., Calteau A., Heidmann T.; |
| RA | "Characterization of a low copy HERV family : evidence for recent |
| RT | integrations in primates of elements with coding envelopes."; |
| RRL | Submitted (AUG-2002) to the EMBL/GenBank/DDJB databases." |
| DR | EMBL; AJ507127; CAD45362.1; - |
| DR | GO; GO:0019028; C:viral capsid; IEA. |
| DR | GO; GO:0005198; F:structural molecule activity; IEA. |
| DR | InterPro; IPR002050; Env_polyprotein. |
| DR | Pfam; PF004429; ENV_polyprotein; 1. |
| SQ | SEQUENCE 584 AA; 65161 MW; DSD52EE2C5494B91 CRC64; |
| QY | Query Match 2.0%; Score 11; DB 6; Length 584; |
| DB | Best Local Similarity 100.0%; Pred. No. 0.058; |
| | Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| RESULT 27 | |
| Q3N2K0 | PRELIMINARY; PRT; 584 AA. |
| ID | Q9N2K0 PRELIMINARY; PRT; 584 AA. |
| AC | Q9N2K0; |
| DT | 01-OCT-2000 (TReMBLrel. 15, Created) |
| DT | 01-OCT-2000 (TReMBLrel. 15, Last sequence update) |
| DT | 01-JUN-2003 (TReMBLrel. 24, Last annotation update) |
| EN | Envelope protein. |
| GN | ENV. |
| OS | HERV-H/env62. |
| OC | Viruses; Retroid viruses; Retroviridae. |
| NCBI_TaxID=129525; | |
| [1] | |
| RN | SEQUENCE FROM N.A. |
| RP | MEDLINE=21109114; PubMed=11162811; |
| RX | de Parseval N., Casella J.P., Gressin L., Heidmann T.; |
| RA | "Characterization of the three envelopes of the HERV-H family |
| RT | encompassing the ISU domain and evolutionary history in primates."; |
| RRL | Virology 279:558-569(2001). |
| DR | EMBL; AJ289709; CAB94192.1; - |
| DR | HSP; P03385; IMOF. |
| DR | GO; GO:0019028; C:viral capsid; IEA. |
| DR | GO; GO:0005198; F:structural molecule activity; IEA. |
| DR | InterPro; IPR002050; Env_polyprotein. |
| DR | Pfam; PF004429; ENV_polyprotein; 2. |
| SQ | SEQUENCE 584 AA; 64318 MW; C03D260B5A60BDAB CRC64; |
| QY | Query Match 2.0%; Score 11; DB 15; Length 584; |
| DB | Best Local Similarity 100.0%; Pred. No. 0.058; |
| | Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |
| RESULT 28 | |
| Q8BI41 | PRELIMINARY; PRT; 618 AA. |
| ID | Q8BI41 PRELIMINARY; PRT; 618 AA. |
| AC | Q8BI41; |
| DT | 01-MAR-2003 (TReMBLrel. 23, Created) |
| DT | 01-MAR-2003 (TReMBLrel. 23, Last sequence update) |

```

RESULT 30
Q9D576
ID Q9D576 PRELIMINARY; PRT; 247 AA.
AC Q9D576;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE 4930505020Rik protein.
GN 4930505020Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=C57BL/6J; TISSUE=Testis;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015711; BAB29942.1; -
DR HSP; P03385; IMCF
DR MGD; MGI:1921959; 4930505020Rik.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env.polyprotein.
DR Pfam; PF00429; ENV.polyprotein; 1.
SQ SEQUENCE 247 AA; 26554 MW; DB9C4F3AA813A5DC CRC64;

Query Match 1.9%; Score 10; DB 11; Length 247;
Best Local Similarity 100.0%; Pred.No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 SLAAVVLQNR 383
DB 199 SLAAVVLQNR 208
|||||
|||||

Search completed: April 19, 2004, 11:03:34
Job time : 50 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 11:02:06 ; Search time 17 seconds
(without alignments)
1647.867 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 538

Sequence: 1 MALPYHIFLFTLLPSFTLT.....PPEETSAQPLLRNSAGSS 538

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 8 | 1.5 | 219 | 1 TPIS PRATU | P96763 francisella |
| 2 | 8 | 1.5 | 497 | 1 T10H TAXCU | Q9axm6 taxus cuspi |
| 3 | 8 | 1.5 | 534 | 1 PSD3 HUMAN | Q43242 homo sapien |
| 4 | 8 | 1.5 | 578 | 1 DSD3 PSESP | Q9k323 pseudomonas |
| 5 | 8 | 1.5 | 669 | 1 MUTL STRAM | Q99uh7 staphylococ |
| 6 | 8 | 1.5 | 669 | 1 MUTL STRAU | Q93c05 staphylococ |
| 7 | 8 | 1.5 | 669 | 1 MUTL STRAU | Q8nw99 staphylococ |
| 8 | 7 | 1.3 | 87 | 1 CYC5 PSEME | P00121 pseudomonas |
| 9 | 7 | 1.3 | 119 | 1 ACPS CHLTR | O84102 chlamydia t |
| 10 | 7 | 1.3 | 119 | 1 ACPS STRA3 | Q8e3m8 streptococ |
| 11 | 7 | 1.3 | 122 | 1 ACPS CHLCV | Q820e7 chlamydothi |
| 12 | 7 | 1.3 | 122 | 1 ACPS CHLPN | Q928m5 chlamydia p |
| 13 | 7 | 1.3 | 125 | 1 ACPS CHLMO | Q9p8t6 chlamydia m |
| 14 | 7 | 1.3 | 127 | 1 ACPS SHEON | Q8eh77 shewanella |
| 15 | 7 | 1.3 | 129 | 1 PURL METTM | P26500 methanobact |
| 16 | 7 | 1.3 | 136 | 1 HNS ECOLI | P08936 escherichia |
| 17 | 7 | 1.3 | 136 | 1 HNS SHIFL | P09120 shigella fl |
| 18 | 7 | 1.3 | 144 | 1 NUS3 THEAC | Q03589 thermoplas |
| 19 | 7 | 1.3 | 183 | 1 ATPF PORPU | P51244 porphyra pu |
| 20 | 7 | 1.3 | 199 | 1 MLO3 SCHPO | Q09330 schizosacch |
| 21 | 7 | 1.3 | 201 | 1 ENV MLVKI | P04502 kirsten mur |
| 22 | 7 | 1.3 | 211 | 1 ENV MSVFB | P03387 fbj murine |
| 23 | 7 | 1.3 | 212 | 1 ER21 BOVIN | P33946 bos taurus |
| 24 | 7 | 1.3 | 212 | 1 ER21 HUMAN | P24390 homo sapien |
| 25 | 7 | 1.3 | 227 | 1 YOR2 EAV | P28992 equine arte |
| 26 | 7 | 1.3 | 240 | 1 RT03 CHOCR | P48938 chondrus cr |
| 27 | 7 | 1.3 | 243 | 1 YC22 METJA | Q58619 methanococ |
| 28 | 7 | 1.3 | 248 | 1 Y738 METTH | O26833 methanococ |
| 29 | 7 | 1.3 | 249 | 1 BTUD ECO57 | Q8x5w0 escherichia |
| 30 | 7 | 1.3 | 249 | 1 BTUD ECO67 | Q8fh28 escherichia |
| 31 | 7 | 1.3 | 249 | 1 BTUD ECOLI | P06611 escherichia |
| 32 | 7 | 1.3 | 249 | 1 BTUD SALTY | Q8xgf8 salmonella |
| 33 | 7 | 1.3 | 251 | 1 RPIA_CABEL | P41394 caenorhabdi |

ALIGNMENTS

| | | | | | | |
|-----|---|-----|-----|---|------------|---------------------|
| 34 | 7 | 1.3 | 273 | 1 | VPRT_ASPB7 | Q00946 african swi |
| 35 | 7 | 1.3 | 273 | 1 | VPRT_ASPW2 | Q85228 african swi |
| 36 | 7 | 1.3 | 280 | 1 | MSAL_SARM2 | Q01416 sarcocystis |
| 37 | 7 | 1.3 | 292 | 1 | Y4AD_RHISN | P53351 rhizobium s |
| 38 | 7 | 1.3 | 297 | 1 | CTAA_BACPF | Q44443 bacillus ps |
| 39 | 7 | 1.3 | 300 | 1 | PYRD_PYRFU | Q8u0p6 pyrococcus |
| 40 | 7 | 1.3 | 303 | 1 | PYRD_PYRAB | Q9v0y6 pyrococcus |
| 41 | 7 | 1.3 | 304 | 1 | NXP4_RAT | Q362n4 rattus norv |
| 42 | 7 | 1.3 | 305 | 1 | MALM_SALTY | P26478 salmonella |
| 43 | 7 | 1.3 | 306 | 1 | MALM_ECOLI | P03841 escherichia |
| 44 | 7 | 1.3 | 308 | 1 | GAAB_PYRAB | Q9v0i7 pyrococcus |
| 45 | 7 | 1.3 | 336 | 1 | PYRD_ECOL6 | Q8fj91 escherichia |
| 46 | 7 | 1.3 | 338 | 1 | PURS_LACLA | Q8ciff4 lactococcus |
| 47 | 7 | 1.3 | 340 | 1 | PURS_LACLC | Q8h186 lactococcus |
| 48 | 7 | 1.3 | 347 | 1 | NU2M_ORNAN | Q36451 ornithorhyn |
| 49 | 7 | 1.3 | 360 | 1 | Y123_BUCAI | P37223 buchera ap |
| 50 | 7 | 1.3 | 364 | 1 | Y115_BUCAP | Q8kal2 buchera ap |
| 51 | 7 | 1.3 | 378 | 1 | POTA_SALTY | P40790 salmonella |
| 52 | 7 | 1.3 | 384 | 1 | KF25_HUMAN | Q9u1l4 homo sapien |
| 53 | 7 | 1.3 | 397 | 1 | CD61_SULSO | Q980n4 sulfolobus |
| 54 | 7 | 1.3 | 398 | 1 | CD61_SULTO | Q975x3 sulfolobus |
| 55 | 7 | 1.3 | 405 | 1 | Y4RN_RHISN | P55647 rhizobium s |
| 56 | 7 | 1.3 | 406 | 1 | CD66_AERPE | Q9yev6 aeropyrum p |
| 57 | 7 | 1.3 | 407 | 1 | PCAB_PSEPU | P32427 pseudomonas |
| 58 | 7 | 1.3 | 414 | 1 | CD61_METAC | Q8tur2 methanosarc |
| 59 | 7 | 1.3 | 414 | 1 | CD61_METMA | Q8pxa8 methanosarc |
| 60 | 7 | 1.3 | 430 | 1 | PBX1_MOUSE | P40424 homo sapien |
| 61 | 7 | 1.3 | 430 | 1 | PBX1_MOUSE | P41778 mus musculu |
| 62 | 7 | 1.3 | 430 | 1 | TOLB_YERPE | Q8zg21 versinia pe |
| 63 | 7 | 1.3 | 431 | 1 | PKNA_WYCTU | P71585 mycobacteri |
| 64 | 7 | 1.3 | 436 | 1 | WNTH_DEIRA | Q9rt08 deinococcus |
| 65 | 7 | 1.3 | 442 | 1 | FTSA_RHIME | Q30994 rhizobium m |
| 66 | 7 | 1.3 | 449 | 1 | STRK_STRGR | P09401 streptomyce |
| 67 | 7 | 1.3 | 451 | 1 | GAGD_DROME | P10405 dirosophila |
| 68 | 7 | 1.3 | 457 | 1 | DLDH_MYCGE | P47513 mycoplasma |
| 69 | 7 | 1.3 | 457 | 1 | YC47_SYNY3 | P45451 synectocyst |
| 70 | 7 | 1.3 | 462 | 1 | GSA_WYCTU | Q06390 mycobacteri |
| 71 | 7 | 1.3 | 468 | 1 | SECV_PYRAB | Q9v1v8 pyrococcus |
| 72 | 7 | 1.3 | 468 | 1 | SECV_PYRFU | Q8u0l3 pyrococcus |
| 73 | 7 | 1.3 | 468 | 1 | SECV_PYRHO | Q94422 pyrococcus |
| 74 | 7 | 1.3 | 473 | 1 | PSBC_ABIAL | Q8b522 abies alba |
| 75 | 7 | 1.3 | 473 | 1 | PSBC_MARPO | P06414 marchantia |
| 76 | 7 | 1.3 | 473 | 1 | PSBC_PINTH | P41643 pinus thunb |
| 77 | 7 | 1.3 | 494 | 1 | SECV_AERPE | Q9ydd0 aeropyrum p |
| 78 | 7 | 1.3 | 510 | 1 | PME2_CITSI | Q04887 citrus sine |
| 79 | 7 | 1.3 | 518 | 1 | CD63_HALN1 | Q9hms3 halobacteri |
| 80 | 7 | 1.3 | 541 | 1 | MLES_OENOE | Q48796 oenococcus |
| 81 | 7 | 1.3 | 563 | 1 | ENV_BAEVM | P10269 baboon endo |
| 82 | 7 | 1.3 | 567 | 1 | ENV_AVISN | P31796 avian splee |
| 83 | 7 | 1.3 | 574 | 1 | ENV_SRV2 | P51515 simian retr |
| 84 | 7 | 1.3 | 574 | 1 | ENV_SMRVH | P21412 squirrel mo |
| 85 | 7 | 1.3 | 575 | 1 | ENV_SMRVH | P21412 squirrel mo |
| 86 | 7 | 1.3 | 582 | 1 | ENV_AVIRE | P03399 avian retic |
| 87 | 7 | 1.3 | 586 | 1 | ENV_MPMV | P07575 simian maso |
| 88 | 7 | 1.3 | 587 | 1 | ENV_SRV1 | P04027 simian retr |
| 89 | 7 | 1.3 | 594 | 1 | DCP2_SCHPO | Q92345 schizosacch |
| 90 | 7 | 1.3 | 616 | 1 | VGF_HUMAN | O15240 homo sapien |
| 91 | 7 | 1.3 | 617 | 1 | VGF_RAT | P20156 rattus norv |
| 92 | 7 | 1.3 | 631 | 1 | FTSK_AZOBK | Q83045 azospirillu |
| 93 | 7 | 1.3 | 633 | 1 | AGP1_YEAST | P25376 saccharomyc |
| 94 | 7 | 1.3 | 636 | 1 | ENV_MCFP | P50073 mink cell f |
| 95 | 7 | 1.3 | 637 | 1 | GYRB_BACSD | O50627 bacillus ha |
| 96 | 7 | 1.3 | 638 | 1 | GYRB_BACSD | P05652 bacillus su |
| 97 | 7 | 1.3 | 639 | 1 | ENV_FLVSA | P06752 feline leuk |
| 98 | 7 | 1.3 | 640 | 1 | ENV_MCFE3 | P03388 mink cell f |
| 99 | 7 | 1.3 | 640 | 1 | ENV_EMCV | P06445 rauscher mi |
| 100 | 7 | 1.3 | 641 | 1 | RAFP_PEPDE | P43466 pedicoccus |

RESULT 1

```

TPIS FRATU
ID _TPIS FRATU STANDARD; PRT; 219 AA.
AC P96763;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).
GN TPIA OR TPI.
OS Francisella tularensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Thiotrichales;
OC Francisellaceae; Francisella.
CX NCBI_TaxID=263;
RN SEQUENCE FROM N.A.
RC STRAIN=LVS;
RX MEDLINE=97188457; PubMed=9037042;
RA Keeling P.J., Doolittle W.F.;
RT "Evidence that eukaryotic triosephosphate isomerase is of alpha-
RT proteobacterial origin."
RL Proc. Natl. Acad. Sci. U.S.A. 94:1270-1275(1997).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glyceralone
CC phosphate.
CC -!- PATHWAY: Plays an important role in several metabolic pathways.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: Belongs to the triosephosphate isomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U73962; AAB48822.1; -.
DR HSSP; P04790; 1TRE.
DR HANAP; MF 00147; -.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRFAMs; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW Pentose shunt.
FT NON_TER 1
FT ACT_SITE 83 BY SIMILARITY.
FT ACT_SITE 155 BY SIMILARITY.
FT NON_TER 219
FT SEQUENCE 219 AA; 23908 MW; 3CD5C967BFF788CC CRC64;
SQ
Query Match 1.5%; Score 8; DB 1; Length 219;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 VKEVISQL 136
DB 36 VKEVISQL 43
RESULT 2
T10H TAXCU STANDARD; PRT; 497 AA.
AC Q9AXM6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Taxane 10-beta-hydroxylase (EC 1.14.13.76) (5-alpha-taxadienol-10-
DE beta-hydroxylase) (Cytochrome P450 725A1).
GN CYP725A1.
OS Taxus cuspidata (Japanese yew).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Taxaceae; Taxus.
CX NCBI_TaxID=99806;

```

```

RN RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=21117067; PubMed=11171980;
RA Schoendorf A., Rithner C.D., Williams R.M., Croteau R.B.;
RT Molecular cloning of a cytochrome P450 taxane 10beta-hydroxylase cDNA
RT from Taxus and functional expression in yeast."
RL Proc. Natl. Acad. Sci. U.S.A. 98:1501-1506(2001).
CC -!- FUNCTION: Involved in the transformation of a taxadienyl acetate-
CC by hydroxylation at C10 to yield taxadien-salphenyl-acetate-10beta-
CC ol.
CC -!- CATALYTIC ACTIVITY: Taxa-4(20),11-dien-5-alpha-yl acetate + NADPH
CC + O(2) = 10-beta-hydroxytaxa-4(20),11-dien-5-alpha-yl acetate +
CC NADP + H(2)O.
CC -!- PATHWAY: Taxol biosynthesis; fourth step.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF318211; AAK00946.1; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00667; P450_1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Taxol biosynthesis; Oxidoreductase; Monooxygenase; Heme; NADP.
FT METAL 443 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT SEQUENCE 497 AA; 56690 MW; CFE40415ABC43814 CRC64;
SQ
Query Match 1.5%; Score 8; DB 1; Length 497;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 457 AIILLILF 464
DB 33 AIILLILF 40
RESULT 3
P5D3_HUMAN STANDARD; PRT; 534 AA.
ID P5D3_HUMAN STANDARD; PRT; 534 AA.
AC O43242; Q96E12; Q9BQA4;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 26S proteasome non-ATPase regulatory subunit 3 (26S proteasome
DE regulatory subunit S3) (Proteasome subunit p58).
GN PSMD3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Hepatoblastoma;
RX MEDLINE=97170075; PubMed=9017604;
RA Kominami K.-I., Okura N., Kawamura M., Demattino G.N., Slaughter C.A.,
RA Shimbara N., Chung C.H., Fujimuro M., Yokosawa H., Shimizu Y.,
RA Tanahashi N., Tanaka K., Toh-E A.;
RT "Yeast counterparts of subunits S5a and p58 (S3) of the human 26S
RT proteasome are encoded by two multicopy suppressors of nin1-1."
RL Mol. Biol. Cell 8:171-187(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye, Lung, Pancreas, and Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.A., Locuallano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Acts as a regulatory subunit of the 26 proteasome which
CC is involved in the ATP-dependent degradation of ubiquitinated
CC proteins.
CC -!- SUBUNIT: The 26S proteasome is composed of a core protease, known
CC as the 20S proteasome, capped at one or both ends by the 19S
CC regulatory complex (RC). The RC is composed of at least 18
CC different subunits in two subcomplexes: the base and the lid,
CC which form the portions proximal and distal to the 20S proteolytic
CC core, respectively.
CC -!- SIMILARITY: Contains 1 PCI domain.
CC -!- SIMILARITY: Belongs to the proteasome subunit S3 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D67025; BAR23651.1; -
CC EMBL; BC000074; AAH00074.1; -
CC EMBL; BC004859; AAH04859.1; -
CC EMBL; BC012302; AAH12302.1; -
CC EMBL; BC020518; AAH20518.1; -
CC EMBL; BC025686; AAH25686.1; -
CC Genew; HGNC:9560; PSM03.
CC GK; O43242; -
CC GO; GO:0005837; C:26S proteasome; TAS.
CC InterPro; IPR000717; PCI.
CC Pfam; PF01399; PCI; 1.
CC SMART; SM00088; PINT; 1.
KW Proteasome.
FT DOMAIN 26 32 POLY-PRO.
FT CONFLICT 60 60 A -> V (IN REF. 1).
FT CONFLICT 157 158 TP -> NT (IN REF. 2; AAH12302).
FT CONFLICT 304 304 E -> V (IN REF. 2; AAH12302).
FT SEQUENCE 534 AA; 60977 MW; 3B3FB5593542C078 CRC64;
SQ
Query Match 1.5%; Score 8; DB 1; Length 534;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 381 QNRRALDL 389
DB 198 QNRRALDL 205
RESULT 4
ID DSD_PESP STANDARD; PRT; 578 AA.
AC Q9KJ23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thiol:disulfide interchange protein dsbd precursor (EC 1.8.1.8)

DE DSDP (Protein-disulfide reductase) (Disulfide reductase).
GN DSDP
OS Pseudomonas sp. (strain JR1 / K1).
OC Bacteria; Proteobacteria.
OX NCBI_TaxID=306;
RN [1]
RP SEQUENCE FROM N.A.
RA Johann A., Michel J., Averhoff B., Gottschalk G.;
RT "The two-component signal transduction system ArmRS is involved in
RT deregulation of 3-isopropylcatechol dioxygenase in the mutant strain
RT Pseudomonas sp. K1";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required to facilitate the formation of correct
CC disulfide bonds in some periplasmic proteins and for the assembly
CC of the periplasmic c-type cytochromes. Acts by transferring
CC electrons from cytoplasmic thioredoxin to the periplasm. This
CC transfer involves a cascade of disulfide bond formation and
CC reduction steps (By similarity).
CC -!- CATALYTIC ACTIVITY: Protein dithiol + NAD(P)+ = protein disulfide
CC + NAD(P)H.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the thioredoxin family. Dsbd subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155506; AAF80267.1; -
CC HAMAP; MF_00399; -; 1.
CC InterPro; IPR003834; Cytococh TM.
CC InterPro; IPR006662; ThioRed.
CC InterPro; IPR006663; Thioredox_dom2.
CC Pfam; PF02683; Dsbd; 1.
CC PROSITE; PS00194; THIOREDOXIN; 1.
KW Oxidoreductase; Redox-active center; Electron transport; NAD;
KW Transmembrane; Inner membrane; Cytochrome c-type biogenesis; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 578 THIO:DISULFIDE INTERCHANGE PROTEIN DSDP.
FT DOMAIN 19 158 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 159 179 POTENTIAL.
FT DOMAIN 180 207 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 208 228 POTENTIAL.
FT DOMAIN 229 237 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 238 258 POTENTIAL.
FT DOMAIN 259 281 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 282 302 POTENTIAL.
FT DOMAIN 303 321 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 322 342 POTENTIAL.
FT DOMAIN 343 351 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 352 372 POTENTIAL.
FT DOMAIN 373 373 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 374 394 POTENTIAL.
FT DOMAIN 395 408 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 409 429 POTENTIAL.
FT DOMAIN 430 478 PERIPLASMIC (POTENTIAL).
FT DISULFID 118 124 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 179 299 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 491 494 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 578 AA; 62181 MW; 7473FC17A9BF7284 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 578;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 326 GAGVLGAL 333
DB 284 GAGVLGAL 291

repair. May act as a "molecular matchmaker", a protein that promotes the formation of a stable complex between two or more DNA-binding proteins in an ATP-dependent manner without itself being part of a final effector complex (By similarity).

!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use...By non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP004826; BAB95044.1; -
 HAMAP; MF_00149; -; 1
 InterPro; IPR003594; ATPbind_ATPase.
 InterPro; IPR002099; DNA_mis_repair.
 Pfam; PF01119; DNA_mis_repair; 1.
 Pfam; PF02518; HATPase_C; 1.
 TIGRFAMs; TIGR00585; mutL; 1.
 PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DNA repair; Complete proteome.

SEQUENCE 689 AA; 76997 MW; 68DB0ED3FBF3C2DD CRC64;

Query Match 1.5%; Score 9; DB 1; Length 669;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 DVNDIKGT 518
 DB 462 DVNDIKGT 469

RESULT 8

CYC5_PSEME STANDARD; PRT; 87 AA.

AC P00121;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DE Cytochrome C5.
 OS Pseudomonas mendocina.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=300;
 RN [1]
 RP SEQUENCE.

RC STRAIN=CH110;
 RA Ambler R.P., Taylor E.;
 RT "Amino acid sequence of cytochrome c-5 from Pseudomonas mendocina.";
 RL Biochem. Sec. Trans. 1:166-168(1973).
 CC !- FUNCTION: THIS IS A MONOHEME C-TYPE CYTOCHROME. IT IS UNREACTIVE WITH CYTOCHROME C REDUCTASE OR OXIDASE BUT SEEMS TO FUNCTION AS AN INTERMEDIATE IN NITRATE RESPIRATION OF FACULTATIVE ANAEROBIC PSEUDOMONADS.
 CC !- SUBUNIT: Homodimer.
 CC PIR; A00144; CCPSYM.
 DR HSP; P11732; ICCS.
 DR InterPro; IPR002323; Cyt_C1E.
 DR InterPro; IPR00345; CytC_heme_BS.
 DR PRINTS; PR00607; CYTOCHROME_CIE.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW Electron transport; Heme.
 FT BINDING 19 19 HEME (COVALENT).
 FT BINDING 22 22 HEME (COVALENT).
 FT METAL 23 23 IRON (HEME AXIAL LIGAND).
 FT METAL 63 63 IRON (HEME AXIAL LIGAND).
 FT DISULFID 69 72 BY SIMILARITY.
 SQ SEQUENCE 87 AA; 8657 MW; 90159AE3FOADE8FA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 87;
 Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 GAGVLGA 332
 DB 24 GAGVLGA 30

RESULT 9

ACPS_CHLTR STANDARD; PRT; 119 AA.

AC O84102;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase) (4'-phosphopantetheinyl transferase acps).
 DE ACPS OR Cti00.
 GN Chlamydia trachomatis.
 OS Chlamydia; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC Bacteria; Chlamydiae;
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lamell C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC !- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme A to a Ser of acyl-carrier protein (By similarity).
 CC !- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine 3',5'-bisphosphate + holo-[acyl-carrier protein].
 CC !- COFACTOR: Magnesium (By similarity).
 CC !- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC !- SIMILARITY: Belongs to the P-pant transferase superfamily. Acps family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB001284; AAC67691.1; -
 PIR; C71556; C71556.
 DR HAMAP; MF_00101; -; 1.
 DR InterPro; IPR008278; 4-PPT_transf.
 DR InterPro; IPR002592; ACPS.
 DR InterPro; IPR004568; Pantethn_trn.
 DR Pfam; PF01648; ACPS; 1.
 DR ProDom; PD004282; ACPS; 1.
 DR TIGRFAMs; TIGR00516; acps; 1.
 DR TIGRFAMs; TIGR00556; pantethn_trn; 1.
 DR Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium; Complete proteome.
 KW METAL 7 7 MAGNESIUM (BY SIMILARITY).
 FT METAL 56 56 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 119 AA; 13274 MW; DCAB558623769CF9 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 119;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 ALGTGIG 338
 DB 61 ALGTGIG 67

RESULT 10

ACPS STRA3 STANDARD; PRT; 119 AA.

AC Q8E3M8; Q8DY09; DB 1; Length 119;
Query Match 1.3%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase acps).
DE ACPS OR GBS1729 OR SAG1685.
GN Streptococcus agalactiae (serotype III), and
OS Streptococcus agalactiae (serotype V).
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495; 216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MEM316 / Serotype III;
RX MEDLINE=2222988; PubMed=12354221;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Deboy L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA Madoff R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carly H.A., Cline R.T., Van Aken S.E., Gill J., Scarcelli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC A to a Ser of acyl-carrier protein (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC 3', 5'-bisphosphate + holo-[acyl-carrier protein].
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the P-pant transferase superfamily. Acps
CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AL766852; CAD47388.1; -;
EMBL; AB014268; RAN00549.1; -;
DR SAGal1st; GBS1729; -;
DR TIGR; SAG1885; -;
DR HAMAP; MF_00101; -; 1.
DR InterPro; IPR008278; 4-PPT_transf.
DR InterPro; IPR002582; ACPS.
DR InterPro; IPR004568; Pantethn_trn.
DR Pfam; PF01648; ACPS; 1.
DR ProDom; PD004282; ACPS; 1.
DR TIGRFAMs; TIGR00516; acps; 1.
DR TIGRFAMs; TIGR00556; pantethn_trn; 1.
KW Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
KW Complete proteome.
FT METAL 8 8 MAGNESIUM (BY SIMILARITY).
FT METAL 59 59 MAGNESIUM (BY SIMILARITY).

SQ SEQUENCE 119 AA; 13268 MW; COEBE77DA27C5C2E CRC64;
Query Match 1.3%; Score 7; DB 1; Length 119;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 332 ALGTGIG 338
DB 64 ALGTGIG 70

RESULT 11
ACPS_CHLCV STANDARD; PRT; 122 AA.
AC Q820E7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase acps).
DE ACPS OR CCA00469.
GN Chlamydomonas reinhardtii.
OS Chlamydomonas reinhardtii.
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiaceae; Chlamydiales.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC.
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapple E., Khouri H., Federova N.B., Carty H.A.,
RA Unayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomonas reinhardtii (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiales.";
RL Nucleic Acids Res. 31:2134-2147(2003).
CC -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC A to a Ser of acyl-carrier protein (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC 3', 5'-bisphosphate + holo-[acyl-carrier protein].
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the P-pant transferase superfamily. Acps
CC family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB016995; AAP05214.1; -;
DR TIGR; CCA00469; -; 1.
DR HAMAP; MF_00101; -; 1.
DR InterPro; IPR008278; 4-PPT_transf.
DR InterPro; IPR002582; ACPS.
DR InterPro; IPR004568; Pantethn_trn.
DR Pfam; PF01648; ACPS; 1.
DR ProDom; PD004282; ACPS; 1.
DR TIGRFAMs; TIGR00516; acps; 1.
DR TIGRFAMs; TIGR00556; pantethn_trn; 1.
KW Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
KW Complete proteome.
FT METAL 9 9 MAGNESIUM (BY SIMILARITY).
FT METAL 58 58 MAGNESIUM (BY SIMILARITY).

SQ SEQUENCE 122 AA; 13722 MW; 4B84923BB20D141C CRC64;
Query Match 1.3%; Score 7; DB 1; Length 122;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 332 ALGTGIG 338
DB 63 ALGTGIG 69

RESULT 12
ACPS_CHLPM STANDARD; PRT; 122 AA.
AC Q928M5; Q9JQ6;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
DE (4'-phosphopantetheinyl transferase acps).
GN ACPS OR CPN0313 OR CP0445 OR CPB0323
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OC NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AK39."
RT Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=TW-183;
RA Geng M.M., Schuhmacher A., Muehlendorfer I., Bensch K.W., Schaefer K.P.,
RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
RA "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
RT other Chlamydia strains based on whole genome sequence analysis."
RL Submitted (May-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
CC A to a Ser of acyl-carrier protein (By similarity).
CC -!- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
CC 3',5'-bisphosphate + holo-[acyl-carrier protein].
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the P-Pant transferase superfamily. Acps
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002304; AAF39234.1; -

```

DR PIR: D81710; D81710.
 DR TIGR: TC0376; -.
 DR HAMAP: MF_00101; -.
 DR InterPro: IPR008278; 4-PPT transf.
 DR InterPro: IPR002582; ACPS.
 DR InterPro: IPR004568; Pantethn_trn.
 DR Pfam: PF01648; ACPS; 1.
 DR ProDom: PD004282; ACPS; 1.
 DR TIGRFAMs: TIGR00516; acps; 1.
 DR TIGRFAMs: TIGR00556; pantethn_trn; 1.
 DR TIGRFAMs: TIGR00556; pantethn; 1.
 KW Transferase; Lipid synthesis; Fatty acid biosynthesis; Magnesium;
 Complete proteome.
 FT METAL 56 7 MAGNESIUM (BY SIMILARITY).
 FT METAL 56 7 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 125 AA; 13897 MW; 833B406F4F8C41A2 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 125;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 332 ALGTGIG 338
 DB 61 ALGTGIG 67
 RESULT 14
 ACPS SHEON STANDARD; PRT; 127 AA.
 AC Q8E177;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Holo-[acyl-carrier protein] synthase (EC 2.7.8.7) (Holo-ACP synthase)
 DE (4'-phosphopantetheinyl transferase acps).
 GN ACPS OR SOI352.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RC MEDLINE=22976986; PubMed=12368613;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
 RA Reddylyum T.V., Smith H.O., Venter J.C., Neelson K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 Shewanella oneidensis.";
 RL Nat. Biotechnol. 20:1118-1123 (2002).
 CC -1- FUNCTION: Transfers the 4'-phosphopantetheine moiety from coenzyme
 A to a Ser of acyl-carrier protein (By similarity).
 CC -1- CATALYTIC ACTIVITY: CoA + apo-[acyl-carrier protein] = adenosine
 3',5'-bisphosphate + holo-[acyl-carrier protein].
 CC -1- COFACTOR: Magnesium (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the P-Pant transferase superfamily. Acps
 family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: AE015579; AN54417.1; .
 DR TIGR: SOI352; .

DR HAMAP: MF_00101; -.
 DR InterPro: IPR008278; 4-PPT transf.
 DR InterPro: IPR002582; ACPS.
 DR InterPro: IPR004568; Pantethn_trn.
 DR Pfam: PF01648; ACPS; 1.
 DR ProDom: PD004282; ACPS; 1.
 DR TIGRFAMs: TIGR00516; acps; 1.
 DR TIGRFAMs: TIGR00556; pantethn_trn; 1.
 DR TIGRFAMs: TIGR00556; pantethn; 1.
 KW Complete proteome.
 FT METAL 58 9 MAGNESIUM (BY SIMILARITY).
 FT METAL 58 9 MAGNESIUM (BY SIMILARITY).
 SQ SEQUENCE 127 AA; 13638 MW; 6C45CC686C50F2A0 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 127;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 332 ALGTGIG 338
 DB 63 ALGTGIG 69
 RESULT 15
 PURL METTM STANDARD; PRT; 129 AA.
 AC P26500;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
 synthase II) (Fragment).
 GN PURL.
 OS Methanobacterium thermoautotrophicum (strain Marburg / DSM 2133).
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=79929;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91244836; PubMed=2037598;
 RA Jenal U., Rechsteiner T., Tan P.-Y., Buehlmann E., Meile L.,
 RA Leisinger T.;
 RT "Isotocetyl-TRNA synthetase of Methanobacterium thermoautotrophicum
 Marburg. Cloning of the gene, nucleotide sequence, and localization
 of a base change conferring resistance to pseudomonadic acid.";
 EL J. Biol. Chem. 266:10570-10577 (1991).
 CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
 ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
 (formamido)-N(1)-(5-phospho-D-ribose)ylacetamide + L-glutamate.
 CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
 CC -1- SUBUNIT: Heterodimer of two subunits, purQ and purL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the FGAMS family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL: M59245; AAA7951.1; .
 DR HAMAP: MF_00420; -.
 DR InterPro: IPR000728; Air_synth.
 DR Pfam: PF00586; AIRS; 1.
 KW Purine biosynthesis; Ligase; ATP-binding.
 FT NP_BIND 91 102
 FT NON_TER 129 129
 SQ SEQUENCE 129 AA; 13951 MW; 2281FC893C20C933 CRC64;
 Query Match 1.3%; Score 7; DB 1; Length 129;
 Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GTGIGGI 340
 DB 93 GTGIGGI 99

RESULT 16
 HNS_ECOLI STANDARD; PRT; 136 AA.

ID HNS_ECOLI
 AC P08936;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA-binding protein H-NS (Histone-like protein HLP-II) (Protein H1)
 DE (Protein B1)
 GN HNS OR HNSA OR DRDX OR OSWZ OR BGLY OR MSYA OR CUR OR PILG OR TOPS OR
 GN B1237 OR C1701 OR Z2013 OR ECS1739.
 OS Escherichia coli,
 OS Escherichia coli O6, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 217992, 83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=88302110; PubMed=2841565;
 RX Pon C.L., Calogero R.A., Gualerzi C.O.;
 RT "Identification, cloning, nucleotide sequence and chromosomal map
 RT location of hns, the structural gene for Escherichia coli DNA-binding
 RT protein H-NS.";
 RL Mol. Gen. Genet. 212:199-202 (1988).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90220876; PubMed=1691451;
 RX Goeransson M., Sonden B., Nilsson P., Dagberg B., Forsman K.,
 RA Emanuelsson K., Uhlin B.E.;
 RT "Transcriptional silencing and thermoregulation of gene expression in
 RT Escherichia coli.";
 RL Nature 344:682-685 (1990).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=88288045; PubMed=3135462;
 RX Falconi M., Gualtieri M.T., la Teana A., Losso M.A., Pon C.L.;
 RA "Proteins from the prokaryotic nucleoid: primary and quaternary
 RT structure of the 15-kD Escherichia coli DNA binding protein H-NS.";
 RL Mol. Microbiol. 2:323-329 (1988).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92165719; PubMed=1337791;
 RX Ueguchi C., Ito K.;
 RT "Multicopy suppression: an approach to understanding intracellular
 RT functioning of the protein export system.";
 RL J. Bacteriol. 174:1454-1461 (1992).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=91117180; PubMed=2177526;
 RX May G., Dersch P., Haardt M., Middendorf A., Bremer E.;
 RA "The osmZ (bglY) gene encodes the DNA-binding protein H-NS (H1a), a
 RT component of the Escherichia coli K12 nucleoid.";
 RL Mol. Gen. Genet. 224:81-90 (1990).
 RN [6]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=97426617; PubMed=9278503;
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).

[7]
 RN SEQUENCE FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:1137-1155 (1996).
 RN [8]
 RN SEQUENCE FROM N.A.
 RP STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 RT of uropathogenic Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 RN [9]
 RN SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533 (2001).
 RN [10]
 RN SEQUENCE FROM N.A.
 RP STRAIN=O157:H7 / RIMD 050952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22 (2001).
 RN [11]
 RN SEQUENCE OF 1-134 FROM N.A.
 RP STRAIN=K12;
 RX MEDLINE=95291445; PubMed=7773397;
 RA Danchin A., Krin E.;
 RT "Filling the gap between hns and adhE in Escherichia coli K12.";
 RL Microbiology 141:959-960 (1995).
 RN [12]
 RN SEQUENCE OF 1-26.
 RP MEDLINE=91212185; PubMed=2020545;
 RX Kajitani M., Ishihama A.;
 RT "Identification and sequence determination of the host factor gene
 RT for bacteriophage Q beta.";
 RL Nucleic Acids Res. 19:1063-1066 (1991).
 RN [13]
 RN SEQUENCE OF 1-20.
 RP MEDLINE=84178480; PubMed=6370250;
 RX Laine B., Sautiere P., Spassky A., Rimsky S.;
 RT "A DNA-binding protein from E. coli isolation, characterization and
 RT its relationship with proteins H1 and B1.";
 RL Biochem. Biophys. Res. Commun. 119:1147-1153 (1984).
 RN [14]
 RN SEQUENCE OF 1-19.
 RP MEDLINE=91115789; PubMed=2126011;
 RX Yamada H., Muramatsu S., Mizuno T.;
 RT "An Escherichia coli protein that preferentially binds to sharply

```

RT RT curved DNA."
RL J. Biochem. 108:420-425 (1990).
RN [15]
RP SEQUENCE OF 1-19.
RC STRAIN=K12;
RX MEDLINE=99085675; PubMed=9868784;
RA Wasinger V.C., Humphrey-Smith I.;
RT "Small genes/gene-products in Escherichia coli K-12.";
RL FEMS Microbiol. Lett. 169:375-382 (1998).
RN [16]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313 (1997).
RN [17]
RP SEQUENCE OF 1-11.
RC STRAIN=K12 / W3110;
RA Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
RA Frutiger S., Paquet N., Wilkins M., Appel R.D., Bairoch A.,
RA Hochstrasser D.F.;
RL Submitted (SEP-1994) to Swiss-Prot.
RN [18]
RP SEQUENCE OF 1-4.
RC STRAIN=K12 / W3110;
RX MEDLINE=98263247; PubMed=9600841;
RA Wilkins M.R., Gastegger A., Tonella L., Ou K., Tyler M.,
RA Sanchez J.-C., Gooley A.A., Walsh B.J., Bairoch A., Appel R.D.,
RA Williams K.L., Hochstrasser D.F.;
RT "Protein identification with N and C-terminal sequence tags in
proteome projects.";
RL J. Mol. Biol. 278:599-608 (1998).
RN [19]
RP FUNCTION.
RX MEDLINE=95020527; PubMed=7934818;
RA Falconi M., Higgins P.N., Spurio R., Pon C.L., Gualerzi C.O.;
RT "Expression of the gene encoding the major bacterial nucleotide
protein H-NS is subject to transcriptional auto-repression.";
RL Mol. Microbiol. 10:273-282 (1993).
RN [20]
RP STRUCTURE BY NMR OF 90-136.
RX MEDLINE=95180433; PubMed=7875316;
RA Shindo H., Iwaki T., Ieda R., Kurumizaka H., Ueguchi C., Mizuno T.,
RA Worioka S., Nakamura H., Kubonisi H.;
RT "Solution structure of the DNA binding domain of a
nucleoid-associated protein, H-NS, from Escherichia coli.";
RL FEBS Lett. 360:125-131 (1995).
CC -1- FUNCTION: H-NS BINDS TIGHTLY TO DS-DNA, INCREASES ITS THERMAL
CC STABILITY AND INHIBITS TRANSCRIPTION. IT ALSO BINDS TO SS-DNA AND
CC RNA BUT WITH A MUCH LOWER AFFINITY. H-NS HAS POSSIBLE HISTONE-LIKE
CC FUNCTION. MAY BE A GLOBAL TRANSCRIPTIONAL REGULATOR THROUGH ITS
CC ABILITY TO BIND TO CURVED DNA SEQUENCES WHICH ARE FOUND IN
CC REGIONS UPSTREAM OF A CERTAIN SUBSET OF PROMOTERS. IT PLAYS A ROLE
CC IN THE THERMAL CONTROL OF PILI PRODUCTION. IT IS SUBJECT TO
CC TRANSCRIPTIONAL AUTO-REPRESSION. IT BINDS PREFERENTIALLY TO THE
CC UPSTREAM REGION OF ITS OWN GENE RECOGNIZING TWO SEGMENTS OF DNA ON
CC BOTH SIDES OF A BEND CENTERED AROUND -150.
CC -1- SUBUNIT: Homodimer.
CC -1- SIMILARITY: Belongs to the histone-like protein H-NS family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
Query Match 1.3%; Score 7; DB 1; Length 136;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 372 LNSLAAY 378
DB 75 LNSLAAY 81

```

```

RESULT 17
HNS_SHIFL
ID_HNS_SHIFL STANDARD; PRT; 136 AA.
AC P09120;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE DNA-binding protein H-NS (Pathogenesis protein kcpA).
GN HNS OR HNSA OR KCPA OR VIRR OR SF1237 OR SF1233.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE OF 28-136 FROM N.A.
RX STRAIN=YSH6200 / Serotype 2a;
MEDLINE=89343624; PubMed=2668687;
RA Yamada M., Sasakawa C., Okada N., Makino S., Yoshikawa M.;
RT "Molecular cloning and characterization of chromosomal virulence
region kcpA of Shigella flexneri.";
RL Mol. Microbiol. 3:207-213 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
MEDLINE=93023838; PubMed=1406252;
RA Homockyj A.E., Tucker S.C., Maurelli A.T.;
RT "Temperature regulation of Shigella virulence: identification of the
repressor gene virR, an analogue of hns, and partial complementation
by tyrosyl transfer RNA (tRNATyr).";
RL Mol. Microbiol. 6:2113-2124 (1992).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=301 / Serotype 2a;
MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G., III, Rose D.J., Darling A.,
RA Mau B., Ferns N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786 (2003).
CC -1- FUNCTION: H-NS BINDS TIGHTLY TO DS-DNA, INCREASES ITS THERMAL
CC STABILITY AND INHIBITS TRANSCRIPTION. IT ALSO BINDS TO SS-DNA AND
CC RNA BUT WITH A MUCH LOWER AFFINITY. H-NS HAS POSSIBLE HISTONE-LIKE
CC FUNCTION. MAY BE A GLOBAL TRANSCRIPTIONAL REGULATOR THROUGH ITS
CC ABILITY TO BIND TO CURVED DNA SEQUENCES WHICH ARE FOUND IN
CC REGIONS UPSTREAM OF A CERTAIN SUBSET OF PROMOTERS. IT PLAYS A ROLE
CC IN THE THERMAL CONTROL OF PILI PRODUCTION. IT IS SUBJECT TO
CC TRANSCRIPTIONAL AUTO-REPRESSION. IT BINDS PREFERENTIALLY TO THE
CC UPSTREAM REGION OF ITS OWN GENE RECOGNIZING TWO SEGMENTS OF DNA ON
CC BOTH SIDES OF A BEND CENTERED AROUND -150 (BY SIMILARITY).
CC -1- FUNCTION: KCPA IS INVOLVED IN PATHOGENESIS. THE PROTEIN IS
CC REQUIRED BY INVADING BACTERIA FOR SPREAD INTO ADJACENT CELLS
CC (COLONIC EPITHELIAL CELLS). IT IS NEEDED TO ESTABLISH KERATO-
CC CONJUNCTIVITIS PROVOCATION (KCP) BY PATHOGEN BACTERIA. IT IS
CC INVOLVED IN THE THERMAL CONTROL OF EPITHELIAL CELL INVASION.
CC -1- SUBUNIT: Homodimer.
CC -1- SIMILARITY: Belongs to the histone-like protein H-NS family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X65848; CAA47322.1; --
 DR EMBL; X13131; CAA31522.1; --
 DR EMBL; AE015150; AAN42850.1; --
 DR EMBL; AE016982; AAP16735.1; --
 DR PIR; S24755; S24755.
 DR HSP; P08936; 1HNR.
 DR InterPro; IPR001801; Histone HNS.
 DR Pfam; PF00815; Histone HNS; 1.
 DR ProDom; PD007337; Histone HNS; 1.
 KW DNA-binding; Transcription regulation; Repressor; Virulence;
 KW Complete proteome.
 FT INIT MET 0
 SQ SEQUENCE 136 AA; 15408 MW; C7A0773C7C83070 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 136;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 372 LNSLAHV 378
 |||||
 DB 75 LNSLAHV 81

RESULT 18
 NUSA THEAC STANDARD; PRT; 144 AA.
 AC Q03589;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE NUSA protein homolog.
 GN TA0393.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=93027268; PubMed=1408839;
 RA Klenk H.-P., Renner O., Schwass V., Zillig W.;
 RT "Nucleotide sequence of the genes encoding the subunits H, B, A' and
 RT A' of the DNA-dependent RNA polymerase and the initiator tRNA from
 RT Thermoplasma acidophilum.";
 RL Nucleic Acids Res. 20:5226-5226 (1992).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Granel W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 RT acidophilum.";
 RL Nature 407:508-513 (2000).
 CC -1- SIMILARITY: Belongs to the nusa family.
 CC -1- SIMILARITY: Contains 1 KH domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X68198; CAA48283.1; --
 DR EMBL; AL445064; CAC11537.1; ALT_INIT.

DR PIR; S26725; S26725.
 DR InterPro; IPR009019; KH prok.
 DR InterPro; IPR004088; KH_type_1.
 DR PROSITE; PS00084; KH_TYPE_1; 1.
 KW Transcription termination; Transcription antitermination; RNA-binding;
 KW Complete proteome.
 FT DOMAIN 101 144 KH.
 SQ SEQUENCE 144 AA; 16776 MW; 9B522D2423448B0D CRC64;

Query Match 1.3%; Score 7; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 KEVISQL 136
 |||||
 DB 54 KEVISQL 60

RESULT 19
 ATPF PORPU STANDARD; PRT; 183 AA.
 ID ATPF PORPU
 AC P51344.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP synthase B chain (EC 3.6.3.14) (Subunit I).
 GN ATPF.
 OS Porphyra purpurea.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
 OX NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Avonport;
 RA Reith M.E., Munholland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 RT genome.";
 RL Plant Mol. Biol. Rep. 13:333-335 (1995).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -1- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- SIMILARITY: Belongs to the ATPase B chain family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U38804; AAC08130.1; --
 DR PIR; S73165; S73165
 DR InterPro; IPR002146; ATPynt_B/B' sub.
 DR Pfam; PF00430; ATP-synt_B; 1.
 KW Hydrogen ion transport; Transmembrane; CF(0); Chloroplast.
 FT TRANSMEM 28 POTENTIAL.
 SQ SEQUENCE 183 AA; 20591 MW; 441F4869530E5060 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 183;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 ILLLLFG 465
 |||||
 DB 37 ILLLLFG 43

RESULT 20


```

MLO3_SCHPO
ID MLO3_SCHPO STANDARD; PRT; 199 AA.
AC Q03330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein mlo3
GN MLO3 OR SPECID7.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4996;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=97128260; PubMed=8972853;
RA Javerzat J.-P., Cranston G., Allshire R.A.;
RT "Fission yeast genes which disrupt mitotic chromosome segregation
RT when overexpressed."
RL Nucleic Acids Res. 24:4676-4683 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hiderg J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Welljeds I., Vantervelds E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu S., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880 (2002).
CC -!- FUNCTION: Not known, interferes with mitotic chromosome
CC segregation when overexpressed.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -!- SIMILARITY: TO YEAST YRA1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; L42551; AA841270.1; -
CC EMBL; Z98270; CAB10980.1; -
CC PIR; T39861; T39861.
CC GeneDB.SPombe; SPECID7.04; -.
CC InterPro; IPR000504; RNA_rec_mot.
CC Pfam; PF00076; rrm; 1.
CC SMART; SM00360; RRM; 1.
CC PROSITE; P550102; RRM; 1.
-----
DR PROSITE; P500030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding. RNA-BINDING (RRM).
FT DOMAIN 55 134
FT DOMAIN 162 180 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 199 AA; 21791 MW; E8544B8DC39A8EF6 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 199;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 371 QLNSLAA 377
DB 138 QLNSLAA 144
-----
RESULT 21
ENV_MLVKI STANDARD; PRT; 201 AA.
AC P04502;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein (Coat polyprotein) [Contains: Knob protein GP70; Spike
DE protein P15E; R protein] (fragment).
GN ENV.
OS Kirsten murine leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11800;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=85014136; PubMed=6091040;
RA Norton J.D., Connor J., Avery R.J.;
RT "Genesis of Kirsten murine sarcoma virus: sequence analysis reveals
RT recombination points and potential leukaemogenic determinant on
RT parental leukaemia virus genome."
RL Nucleic Acids Res. 12:6839-6852 (1984).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X00982; CAA25490.1; -
CC PIR; A03985; A03985.
CC HSP; P03385; IMOF.
CC InterPro; IPR002050; Env_polyprotein.
CC Pfam; PF00429; Env_polyprotein; 1.
CC Coat protein; Glycoprotein; Polyprotein.
KW NON_TER 1 1
FT CHAIN <1 7 KNOB PROTEIN GP70.
FT CHAIN 8 187 SPIKE PROTEIN P15E.
FT CHAIN 188 201 R PROTEIN.
SQ SEQUENCE 201 AA; 22389 MW; 3A900F7E3A0A3B15 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 201;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 378 VVLQNER 384
DB 74 VVLQNER 80
-----
RESULT 22
ENV_MSVFB STANDARD; PRT; 211 AA.
AC P03387;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein (Coat polyprotein) [Contains: Knob protein GP70; Spike
 GN protein P158; R protein] (Fragment).
 GN ENV.
 OS FSJ murine osteosarcoma virus.
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11805;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83180421; PubMed=6301687;
 RA Van Beveren C., van Straaten F., Curran T., Mueller R., Verma I.M.;
 RT "Analysis of FSJ-MuSV provirus and c-fos (mouse) gene reveals that
 RT viral and cellular fos gene products have different carboxy
 RT termini."
 RL Cell 32:1241-1255(1983).
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; V01184; CAA24506.1; -;
 DR PIR; A03986; A03986.
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; ENV_polyprotein; 1.
 KW Coat protein; glycoprotein; Polyprotein.
 FT NON_TER 1
 FT CHAIN <1 10 KNOB PROTEIN GP70.
 FT CHAIN 11 190 SPIKE PROTEIN P158.
 FT CHAIN 191 211 R PROTEIN.
 FT CHAIN 211 211
 SQ SEQUENCE 211 AA; 23708 MW; 082EC960E0E185B CRC64;

 Query Match 1.3%; Score 7; DB 1; Length 211;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 378 VVLQNR 384
 DB 77 VVLQNR 83

 RESULT 23
 ER21_BOVIN
 ID ER21_BOVIN STANDARD; PRT; 212 AA.
 AC P33946;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ER lumen protein retaining receptor 1 (KDEL receptor 1) (P23).
 GN KDELRI OR KDEL.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93132065; PubMed=8380600;
 RA Tang B.L., Wong S.H., Qi X.L., Low S.H., Hong W.;
 RT "Molecular cloning, characterization, subcellular localization and
 RT dynamics of p23, the mammalian KDEL receptor."
 RL J. Cell Biol. 120:325-338(1993).
 CC -1- FUNCTION: Required for the retention of luminal endoplasmic
 CC reticulum proteins. Determines the specificity of the luminal ER
 CC protein retention system. Also required for normal vesicular
 CC traffic through the Golgi. This receptor recognizes K-D-B-L.
 CC -1- SUBUNIT: Interacts with ARFGAP1 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: Belongs to the ERD2 family.
 DR PIR; A44394; A44394.
 DR InterPro; IPR000133; ERret_receptor.
 DR Pfam; PF00810; ER_lumen_recept; 1.
 DR PRINIS; PRO0860; ERLUMENR.
 DR ProDom; PD005774; ERret_receptor; 1.
 DR PROSITE; PS00951; ER_LUMEN_RECEPTOR_1; 1.
 DR PROSITE; PS00952; ER_LUMEN_RECEPTOR_2; 1.
 KW Endoplasmic reticulum; Transmembrane; Protein transport; Receptor.
 FT DOMAIN 1 2 LUMENAL (POTENTIAL).
 FT TRANSMEM 3 21 POTENTIAL.
 FT DOMAIN 22 35 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 36 53 POTENTIAL.
 FT DOMAIN 54 61 LUMENAL (POTENTIAL).
 FT TRANSMEM 62 80 POTENTIAL.
 FT DOMAIN 81 96 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 97 110 POTENTIAL.
 FT DOMAIN 111 117 LUMENAL (POTENTIAL).
 FT TRANSMEM 118 137 POTENTIAL.
 FT DOMAIN 138 149 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 150 168 POTENTIAL.
 FT DOMAIN 169 178 LUMENAL (POTENTIAL).
 FT TRANSMEM 179 199 POTENTIAL.
 FT DOMAIN 200 212 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 212 AA; 24556 MW; 007FA62BF14AC920 CRC64;

 Query Match 1.3%; Score 7; DB 1; Length 212;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 457 AIIILL 463
 DB 15 AIIILL 21

 RESULT 24
 ER21_HUMAN
 ID ER21_HUMAN STANDARD; PRT; 212 AA.
 AC P24390;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ER lumen protein retaining receptor 1 (KDEL receptor 1).
 GN KDELRI OR ERD2.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=91043069; PubMed=2172835;
 RA Lewis M.J., Pelham H.R.B.;
 RT "A human homologue of the yeast HDEL receptor."
 RL Nature 348:162-163(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko M., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uscin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.N., Krzywinski S.J., Skalska U., Smailus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[3]

MUTAGENESIS.

RP MEDLINE=93327771; PubMed=8392934;

RA Townsley F.M., Wilson D.W., Pelham H.R.B.;

RA "Mutational analysis of the human KDEL receptor: distinct structural requirements for Golgi retention, ligand binding and retrograde transport."; EMBO J. 12:2821-2829(1993).

CC -!- FUNCTION: Required for the retention of luminal endoplasmic reticulum proteins. Determines the specificity of the luminal ER protein retention system. Also required for normal vesicular traffic through the Golgi. This receptor recognizes K-D-E-L.

CC -!- SUBUNIT: Interacts with ARF GAP1 (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- SIMILARITY: Belongs to the ERD2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

EMBL; X55885; CAA39371.1; -.

DR EMBL; BC018778; AAH18778.1; -.

DR PIR; S13293; S13293.

DR GeneW; HGNC:6304; KDELR1.

DR MM; 131235; -.

DR GO; GO:0016021; C: integral to membrane; TAS.

DR GO; GO:0005624; C: membrane fraction; TAS.

DR GO; GO:0005046; F: KDEL sequence binding; TAS.

DR GO; GO:0006886; P: intracellular protein transport; TAS.

DR InterPro; IPR000133; Eret_receptor.

DR Pfam; PF00810; ER_lumen_recept; 1.

DR PRINTS; PRO0660; ERLUMENR.

DR PRODOM; PD005774; Eret receptor; 1.

DR PROSITE; PS00951; ER_LUMEN_RECEPTOR_1; 1.

DR PROSITE; PS00952; ER_LUMEN_RECEPTOR_2; 1.

KW Endoplasmic reticulum; Transmembrane; Protein transport; Receptor.

FT DOMAIN 1 2 LUMENAL (POTENTIAL).

FT TRANSMEM 3 21 POTENTIAL.

FT DOMAIN 22 35 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 36 53 POTENTIAL.

FT DOMAIN 54 61 LUMENAL (POTENTIAL).

FT TRANSMEM 62 80 POTENTIAL.

FT DOMAIN 81 96 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 97 110 POTENTIAL.

FT DOMAIN 111 117 LUMENAL (POTENTIAL).

FT TRANSMEM 118 137 POTENTIAL.

FT DOMAIN 138 149 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 150 168 POTENTIAL.

FT DOMAIN 169 178 LUMENAL (POTENTIAL).

FT TRANSMEM 179 199 POTENTIAL.

FT DOMAIN 200 212 CYTOPLASMIC (POTENTIAL).

SEQ SEQUENCE 212 AA; 24542 MW; B863C0F09CFC0551 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 212;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 AIIIIII 463

DB 15 AIIIIII 21

RESULT 25

YOR2_EAV

ID YOR2_EAV STANDARD; PRT; 227 AA.

AC P28992;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DE Hypothetical 25.6 kDa protein (ORF2).

OS Equine arteritis virus (EAV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Arteriviridae; Arterivirus.

OX NCBI_TaxID=11047;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Eueryus;

RX MEDLINE=91237805; PubMed=1951863;

RA den Boon J.A., Sijder E.J., Chirnside E.D., de Vries A.A.F., Horzinek M.C., Spaan W.J.M.;

RA "Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily.";

RT Coronaviridae; Arterivirus.

RL J. Virol. 65:2910-2920(1991).

CC -!- FUNCTION: Possible envelope protein.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

EMBL; X53459; CAA37541.1; -.

DR PIR; C39925; C39925.

DR InterPro; IPR001913; EAV env prot.

DR Pfam; PF01309; EAV env_prot; 1.

DR PRODOM; PD002461; EAV_env_prot; 1.

KW Hypothetical protein; Envelope protein.

FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).

SEQ SEQUENCE 227 AA; 25579 MW; 6560BEA76C06D0B3 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 227;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 455 LAAILLL 461

DB 186 LAAILLL 192

RESULT 26

RT03_CHOCH

ID RT03_CHOCH STANDARD; PRT; 240 AA.

AC P48938;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Mitochondrial ribosomal protein S3.

GN RPS3.

OS Chondrus crispus (Carragheen).

OG Mitochondrion.

OC Eukaryota; Rhodophyta; Florideophyceae; Gigartinales; Gigartineae;

OC Chondrus.

OX NCBI_TaxID=2769;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE=Apices;

RX MEDLINE=95341681; PubMed=7616569;

RA Leblanc C., Boyen C., Richard O., Bonnard G., Grienberger J.M., Kloareg B.;

RA "Complete sequence of the mitochondrial DNA of the rhodophyte Chondrus crispus (Gigartinales). Gene content and genome organization.";

RT Chondrus crispus (Gigartinales).

RL J. Mol. Biol. 250:484-495(1995).

CC -!- SUBCELLULAR LOCATION: Mitochondrial.

CC -!- SIMILARITY: Belongs to the S3P family of ribosomal proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: Z47547; CA87601.1; -;
 DR PIR: S59085; S59085.
 DR InterPro: IPR009019; KH prok.
 DR InterPro: IPR001351; Ribosomal_S3_C.
 DR InterPro: IPR008282; Ribosomal_S3_N.
 DR Pfam: PF00189; Ribosomal_S3_C; 1.
 DR Pfam: PF00417; Ribosomal_S3_N; 1.
 DR PROSITE: PS00548; RIBOSOMAL_S3; FALSE_NEG.
 KW Ribosomal protein, Mitochondrion.
 SQ SEQUENCE 240 AA; 28142 MW; C910AA2B5D8C800E CRC64;

Query Match 1.3%; Score 7; DB 1; Length 240;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 368 LODOLNS 374
 |||||
 Db 146 LODOLNS 152

RESULT 27

YC22_METJJA STANDARD; PRT; 243 AA.

AC Q58619;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1222.
 GN MJ1222.

OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

EX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gockayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U67563; AAB99226.1; -;
 DR PIR: E64452; E64452.

DR TIGR: MJ1222; -;

DR InterPro: IPR001173; Glyco_trans_2.

DR Pfam: PF00535; Glycosyl_transf_2; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 243 AA; 27428 MW; 23BE4D5026EAS44 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 243;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 GALGTGI 337

Db 80 GALGTGI 86

RESULT 28

Y738_METTH

ID Y738_METTH STANDARD; PRT; 248 AA.

AC Q26833;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hypothetical protein MTH738.
 GN MTH738.

OS Methanobacterium thermoautotrophicum.

OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

OC Methanobacteriaceae; Methanothermobacter.

OX NCBI_TaxID=187420;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Delta H.

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., Blakely D., Cook R., Gilbert K.,
 RA Aldredge T., Bashirzadeh R., Keagle P., Lum W., Pochier B., Qiu D.,
 RA Harrison D., Hoang L., Wang Y., Wierzbowski J., Gibson R.,
 RA Sadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiawani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 DeltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155 (1997).

CC -!- SIMILARITY: TO M.JANNASCHII MJ1452.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: A3000853; AAB85242.1; -;

DR PIR: F69198; F69198.

DR InterPro: IPR000051; SAM bind.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 248 AA; 27157 MW; 57F1D61E7FA6A68 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 248;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 SFTLTAP 22

Db 195 SFTLTAP 201

RESULT 29

BTUD_ECO57

ID BTUD_ECO57 STANDARD; PRT; 249 AA.

AC Q8X5W0;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Vitamin B12 transport ATP-binding protein btud (EC 3.6.3.33)

DE (Vitamin B12-transporting ATPase).

GN BTUD OR Z2738 OR ECS2416.

Search completed: April 19, 2004, 11:07:50
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:55:05 ; Search time 59 Seconds

(without alignments)
2576.450 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 538

Sequence: 1 MALPVIHFLFTVLLPSFTLT.....PPEISAAQPLLRNSAGSS 538

Scoring table: OLIGO Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : A_Geneseq_29Jan04.*

- 1: Geneseq1980a.*
- 2: Geneseq1990a.*
- 3: Geneseq2000a.*
- 4: Geneseq2001a.*
- 5: Geneseq2002a.*
- 6: Geneseq2003a.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004a.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 538 | 100.0 | 538 | 4 | AAB75138 |
| 2 | 538 | 100.0 | 538 | 4 | AAB67652 |
| 3 | 538 | 100.0 | 540 | 3 | ABP40980 |
| 4 | 538 | 100.0 | 685 | 3 | ABP40988 |
| 5 | 538 | 100.0 | 846 | 3 | ABP41050 |
| 6 | 437 | 81.2 | 538 | 2 | AAV08622 |
| 7 | 437 | 81.2 | 538 | 3 | AAV67313 |
| 8 | 437 | 81.2 | 538 | 5 | AAE14540 |
| 9 | 437 | 81.2 | 538 | 7 | ADC38777 |
| 10 | 151 | 29.9 | 544 | 7 | ADS09243 |
| 11 | 153 | 28.4 | 538 | 2 | AAW99345 |
| 12 | 150 | 22.3 | 120 | 4 | ABG75137 |
| 13 | 74 | 13.8 | 446 | 4 | ABG07532 |
| 14 | 74 | 13.8 | 446 | 4 | ABG20007 |
| 15 | 67 | 12.5 | 67 | 2 | AAW42030 |
| 16 | 67 | 12.5 | 67 | 2 | AAW44290 |
| 17 | 52 | 9.7 | 52 | 2 | AAW99346 |
| 18 | 45 | 8.4 | 493 | 2 | AAW71068 |
| 19 | 45 | 8.4 | 493 | 2 | AAW99353 |
| 20 | 45 | 8.4 | 541 | 3 | AAW08195 |
| 21 | 45 | 8.4 | 542 | 3 | AAW19069 |
| 22 | 42 | 7.8 | 70 | 7 | ADD26900 |
| 23 | 42 | 7.8 | 144 | 4 | ABG08534 |
| 24 | 39 | 7.2 | 58 | 4 | ABG24204 |
| 25 | 35 | 6.5 | 493 | 7 | ADB84401 |

oligoma search

| | | | | | | |
|----|-----|------|---|----------|----------|-----------|
| 26 | 5.9 | 32 | 3 | ABP41044 | ABP41044 | HERV-7q C |
| 27 | 5.9 | 542 | 5 | AAE25054 | AAE25054 | Human EMB |
| 28 | 5.8 | 388 | 4 | ABG28311 | ABG28311 | Novel hum |
| 29 | 5.8 | 633 | 4 | ABG28306 | ABG28306 | Novel hum |
| 30 | 5.4 | 29 | 3 | ABP41042 | ABP41042 | Human ret |
| 31 | 5.4 | 31 | 4 | ABG01777 | ABG01777 | Novel hum |
| 32 | 5.4 | 31 | 5 | ABP63036 | ABP63036 | Human pol |
| 33 | 4.6 | 263 | 4 | AAW18008 | AAW18008 | Peptide # |
| 34 | 4.6 | 263 | 4 | ABP37043 | ABP37043 | Peptide # |
| 35 | 4.6 | 263 | 4 | AAW30519 | AAW30519 | Peptide # |
| 36 | 4.6 | 263 | 4 | ABP31812 | ABP31812 | Peptide # |
| 37 | 4.6 | 263 | 4 | ABP22358 | ABP22358 | Protein # |
| 38 | 4.6 | 263 | 4 | AAW70185 | AAW70185 | Human bon |
| 39 | 4.6 | 263 | 4 | AAW57770 | AAW57770 | Human bra |
| 40 | 4.6 | 263 | 4 | ABG51885 | ABG51885 | Human liv |
| 41 | 4.6 | 263 | 4 | AAW05648 | AAW05648 | Peptide # |
| 42 | 4.6 | 263 | 5 | ABG39819 | ABG39819 | Human pep |
| 43 | 3.7 | 20 | 4 | AAW75139 | AAW75139 | HERV-W en |
| 44 | 3.7 | 20 | 4 | ABP41043 | ABP41043 | HE2 CKS-1 |
| 45 | 3.2 | 303 | 4 | ABG05327 | ABG05327 | Novel hum |
| 46 | 3.0 | 157 | 4 | ABG07530 | ABG07530 | Novel hum |
| 47 | 3.0 | 157 | 4 | ABG20006 | ABG20006 | Novel hum |
| 48 | 2.4 | 87 | 4 | ABG15631 | ABG15631 | Novel hum |
| 49 | 2.4 | 87 | 4 | ABG27347 | ABG27347 | Novel hum |
| 50 | 2.4 | 162 | 2 | AAW71069 | AAW71069 | Multiple |
| 51 | 2.4 | 162 | 2 | AAW99554 | AAW99554 | Protein e |
| 52 | 2.4 | 162 | 7 | ADB84404 | ADB84404 | MSRV-1 as |
| 53 | 2.4 | 426 | 4 | ABG08531 | ABG08531 | Novel hum |
| 54 | 2.2 | 20 | 4 | AAW75141 | AAW75141 | HERV-W en |
| 55 | 2.2 | 36 | 4 | ABP41410 | ABP41410 | Peptide # |
| 56 | 2.2 | 36 | 4 | AAW35202 | AAW35202 | Peptide # |
| 57 | 2.2 | 36 | 4 | ABP25326 | ABP25326 | Protein # |
| 58 | 2.2 | 36 | 4 | AAW75084 | AAW75084 | Human bon |
| 59 | 2.2 | 36 | 4 | ABG62280 | ABG62280 | Human bra |
| 60 | 2.2 | 36 | 4 | ABG56848 | ABG56848 | Human liv |
| 61 | 2.2 | 36 | 5 | ABG44781 | ABG44781 | Human pep |
| 62 | 2.2 | 54 | 5 | ABP64532 | ABP64532 | Human ORF |
| 63 | 2.0 | 63 | 4 | ABP41872 | ABP41872 | Peptide # |
| 64 | 2.0 | 63 | 4 | AAW35671 | AAW35671 | Peptide # |
| 65 | 2.0 | 63 | 4 | ABP25570 | ABP25570 | Protein # |
| 66 | 2.0 | 63 | 4 | AAW75564 | AAW75564 | Human bon |
| 67 | 2.0 | 63 | 4 | AAW62746 | AAW62746 | Human bra |
| 68 | 2.0 | 63 | 4 | ABG57306 | ABG57306 | Human liv |
| 69 | 2.0 | 63 | 5 | ABG45099 | ABG45099 | Human pep |
| 70 | 2.0 | 131 | 4 | AAW22050 | AAW22050 | Peptide # |
| 71 | 2.0 | 131 | 4 | ABP44443 | ABP44443 | Peptide # |
| 72 | 2.0 | 131 | 4 | AAW38431 | AAW38431 | Peptide # |
| 73 | 2.0 | 144 | 4 | AAW93523 | AAW93523 | Human pol |
| 74 | 2.0 | 144 | 5 | ABP89170 | ABP89170 | Human pol |
| 75 | 2.0 | 253 | 4 | AAW94312 | AAW94312 | Human rep |
| 76 | 2.0 | 281 | 4 | ABG19856 | ABG19856 | Novel hum |
| 77 | 2.0 | 305 | 4 | ABG22875 | ABG22875 | Novel hum |
| 78 | 2.0 | 356 | 4 | ABG19862 | ABG19862 | Novel hum |
| 79 | 2.0 | 459 | 5 | ABU03726 | ABU03726 | Human ova |
| 80 | 2.0 | 584 | 5 | ABP69150 | ABP69150 | Human pol |
| 81 | 2.0 | 584 | 4 | ABG22871 | ABG22871 | Novel hum |
| 82 | 2.0 | 584 | 4 | ABG19851 | ABG19851 | Novel hum |
| 83 | 2.0 | 584 | 4 | ABG05606 | ABG05606 | Novel hum |
| 84 | 2.0 | 586 | 4 | ABG19838 | ABG19838 | Novel hum |
| 85 | 2.0 | 795 | 4 | ABG16649 | ABG16649 | Novel hum |
| 86 | 2.0 | 1167 | 4 | ABG05605 | ABG05605 | Novel hum |
| 87 | 2.0 | 1186 | 4 | ABG06819 | ABG06819 | Novel hum |
| 88 | 2.0 | 1340 | 4 | ABG06717 | ABG06717 | Novel hum |
| 89 | 1.9 | 10 | 3 | ABP40994 | ABP40994 | Human HER |
| 90 | 1.9 | 10 | 3 | ABP40998 | ABP40998 | Human HER |
| 91 | 1.9 | 10 | 3 | ABP41003 | ABP41003 | Human HER |
| 92 | 1.9 | 10 | 3 | ABP41026 | ABP41026 | Human HER |
| 93 | 1.9 | 10 | 3 | ABP40993 | ABP40993 | Human HER |
| 94 | 1.9 | 10 | 3 | ABP41014 | ABP41014 | Human HER |
| 95 | 1.9 | 10 | 3 | ABP41001 | ABP41001 | Human HER |
| 96 | 1.9 | 10 | 3 | ABP41025 | ABP41025 | Human HER |
| 97 | 1.9 | 10 | 3 | ABP40999 | ABP40999 | Human HER |
| 98 | 1.9 | 10 | 3 | ABP41002 | ABP41002 | Human HER |

99 10 1-9 10 3 ABP41015 Human HER
100 10 1-9 10 3 ABP41016 Human HER

ALIGNMENTS

RESULT 1

AA875138
ID AA875138 standard; protein; 538 AA.

XX AA875138;

DT 08-AUG-2001 (first entry)

DE HERV-W envelope protein G.

XX Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW envelope protein; multiple sclerosis-related superantigen; vaccine;
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW antisense-therapy; autoimmune disorder.

XX Human endogenous retrovirus.

XX Key Location/Qualifiers
FH Region 271..317
FT /label= SU
FT /note= "surface protein"

FT Cleavage-site 317..318

FT /note= "SU-TM putative cleavage site"

FT Region 318..538

FT /label= TM

FT /note= "transmembrane subunit"

XX WO200131021-A1.

XX 03-MAY-2001.

XX 30-OCT-2000; 2000WO-EP010659.

XX 28-OCT-1999; 99EP-00402690.

XX (UYGE-) UNIV GENEVE.

XX Conrad B, Mach B;

XX WPI; 2001-316336/33.

XX N-PSDB; AAH20070.

XX New human retrovirus HERV-W ENV proteins/peptides having superantigen

XX activity useful for diagnosing and treating multiple sclerosis.

XX Claim 1; Fig 7; 94pp; English.

XX On the basis of the PBS t-RNA motif used for the classification of human
XX endogenous retrovirus (HERVs) the full length endogenous provirus which
XX was located on the long arm of human chromosome 7 (7q21-22) has been
XX designated HERV-W. The present invention describes proteins or peptides
XX (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
XX of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
XX have neuroprotective activity, and can be used in: vaccines; antisense-
XX therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
XX useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
XX disorders. (I) are also useful for identifying substances (and optionally
XX recovering) capable of binding to a retroviral superantigen associated
XX with MS, substances capable of blocking SAG activity and substances
XX capable of blocking transcription or translation of HERV-W retroviral
XX superantigen. A protein or peptide derived from (I), modified to be
XX devoid of SAG activity and being capable of generating an immune response
XX against HERV-W retroviral SAG is useful in therapy. Nucleic acid
XX molecules encoding (I) are useful as vaccines against MS. Substances
XX capable of blocking SAG activity, capable of binding to a retroviral
XX superantigen associated with MS, or capable of blocking transcription or

CC translation of HERV-W retroviral superantigen for use in treating or
CC preventing MS obtained using (I) are useful for the treatment and
CC prevention of MS. (I) and nucleic acids encoding them are useful for
CC diagnosing autoimmune disease. The present sequence represents the
CC specifically claimed envelope protein of HERV-W designated G

XX Sequence 538 AA;

Query Match 100.0%; Score 538; DB 4; Length 538;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSPFTLTAPPCCRCMTSSSPYQEFILWRMQRGNIDAPSYLSKQTP 60

DB 1 MALPYHIFLFTVLLPSPFTLTAPPCCRCMTSSSPYQEFILWRMQRGNIDAPSYLSKQTP 60

QY 61 TFTAHTMPCRYHSATLCHMANTHYTGMINSCPGGLGTVTCWYFTCTGMSDGGV 120

DB 61 TFTAHTMPCRYHSATLCHMANTHYTGMINSCPGGLGTVTCWYFTCTGMSDGGV 120

QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLRLVSLFNTLLTGLHEVSA 180

DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLRLVSLFNTLLTGLHEVSA 180

QY 181 QNPTNCWICLPLNFRPYVISIPVPEQWNNFSTEINTTSVLGPLYSNLEITHSLTCKVF 240

DB 181 QNPTNCWICLPLNFRPYVISIPVPEQWNNFSTEINTTSVLGPLYSNLEITHSLTCKVF 240

QY 241 SNTTNTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCCLNGSSSCMCFSLFVPPMTIY 300

DB 241 SNTTNTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCCLNGSSSCMCFSLFVPPMTIY 300

QY 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGGITTSQFYKLSQELNGDMER 360

DB 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGGITTSQFYKLSQELNGDMER 360

QY 361 VADSLVTLDQDLSLAALVQLNRRALDLTAERGCTCLFGECCYVYVQSGIVTEKYKE 420

DB 361 VADSLVTLDQDLSLAALVQLNRRALDLTAERGCTCLFGECCYVYVQSGIVTEKYKE 420

QY 421 IRDRIQREARELNTGPMGLLSQMPWILPFLGPIAAIILLLPGPCIFNLLVNFVSRI 480

DB 421 IRDRIQREARELNTGPMGLLSQMPWILPFLGPIAAIILLLPGPCIFNLLVNFVSRI 480

QY 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLPNSAGSS 538

DB 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLPNSAGSS 538

RESULT 2

AA867652

ID AA867652 standard; protein; 538 AA.

XX AC AA867652;

XX DT 29-MAY-2001 (first entry)

XX Amino acid sequence of a human endogenous retrovirus envelope protein.

XX Envelope protein; HERV; syncytia formation; placental development;

XX syncytia; cancer; cell adhesion.

XX Human endogenous retrovirus.

XX WO200116171-A1.

XX PD 08-MAR-2001.

XX PF 01-SEP-2000; 2000WO-FR002429.

XX PR 01-SEP-1999; 99FR-00011141.

XX PR 15-SEP-1999; 99FR-00011793.

PA (INRM) BIO MERIEUX.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
 XX WPI; 2001-226676/23.
 DR N-PSDB; AAF55630.
 XX
 PT Detecting expression of human endogenous retrovirus envelope protein in
 PT cells of a tissue or culture, from its ability to induce syncytia.
 XX
 XX
 PS Claim 1; Page 41-44; 57pp; French.
 XX
 CC The present sequence represents a human endogenous retrovirus envelope
 CC protein. The specification describes a method for detecting expression of
 CC an envelope protein from a human endogenous retrovirus (HERV), in cells,
 CC of a tissue or culture. The method comprises detecting syncytia formation
 CC due to the fusogenic properties of the envelope protein. Envelope
 CC polypeptides and polynucleotides are used to produce therapeutic or
 CC prophylactic compositions, particularly for treatment of cancer, to
 CC correct defects in placental development (or other natural formation of
 CC other types of syncytia), and to promote adhesion of cells in grafts or
 CC cellular repair processes. Expression of sequences antisense to the
 CC polynucleotide are used to prevent formation of syncytia.
 XX
 XX Sequence 538 AA;
 SQ
 Query Match 100.0%; Score 538; DB 4; Length 538;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMORPCGNIDAPSYRSLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMORPCGNIDAPSYRSLSKGTP 60
 QY 61 TFTAHTMPRCNCHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTVTTQTGMSDGGGV 120
 DB 61 TFTAHTMPRCNCHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTVTTQTGMSDGGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLTVLSLENTTLTGLHEVSA 180
 DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLTVLSLENTTLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHSNLTCKVF 240
 DB 181 QNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHSNLTCKVF 240
 QY 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSNCFLSLVPPMTIY 300
 DB 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSNCFLSLVPPMTIY 300
 QY 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTTQFYKLSQELNGDMER 360
 DB 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTTQFYKLSQELNGDMER 360
 QY 361 VADSLVTLQDLSLAVALVQNRALDILLTAERGCTCLFGECCCYVNVQSGIVTEKVE 420
 DB 361 VADSLVTLQDLSLAVALVQNRALDILLTAERGCTCLFGECCCYVNVQSGIVTEKVE 420
 QY 421 IRDRIQRAEELNRTGPGWLLSOMMPWILFPLGLAAIILLLLFGPCIENLNVFVSSRI 480
 DB 421 IRDRIQRAEELNRTGPGWLLSOMMPWILFPLGLAAIILLLLFGPCIENLNVFVSSRI 480
 QY 481 EAVKLQMEPKWQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRNSAGSS 538
 DB 481 EAVKLQMEPKWQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRNSAGSS 538

RESULT 3

ABP40980

*ID ABP40980 standard; protein; 540 AA.

XX

AC ABP40980;

XX
 DT 01-AUG-2002 (first entry)
 XX
 DE Human retroviral HERV-7q env peptide #3.
 XX
 KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis.
 XX
 OS Human endogenous retrovirus.
 XX
 PN WO9967395-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 23-JUN-1999; 99WO-FR001513.
 XX
 PR 23-JUN-1998; 98FR-00007920.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 XX Alliel PM, Perin J, Rieger F;
 XX WPI; 2000-160587/14.
 DR
 PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 PT for diagnosis, treatment and prevention of autoimmune and neurological
 PT diseases.
 XX
 PS Claim 22; Fig 4; 225pp; French.
 XX
 CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention
 XX
 SQ Sequence 540 AA;
 Query Match 100.0%; Score 538; DB 3; Length 540;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMORPCGNIDAPSYRSLSKGTP 60
 DB 3 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMORPCGNIDAPSYRSLSKGTP 62
 QY 61 TFTAHTMPRCNCHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTVTTQTGMSDGGGV 120
 DB 63 TFTAHTMPRCNCHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTVTTQTGMSDGGGV 122
 QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLTVLSLENTTLTGLHEVSA 180
 DB 123 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTLTVLSLENTTLTGLHEVSA 182
 QY 181 QNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHSNLTCKVF 240
 DB 183 QNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGLPLVSNLEITHSNLTCKVF 242
 QY 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSNCFLSLVPPMTIY 300
 DB 243 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSNCFLSLVPPMTIY 302
 QY 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTTQFYKLSQELNGDMER 360
 DB 303 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTTQFYKLSQELNGDMER 362

QY 361 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCFLGEECCYYVNSGIVTEKYKE 420
Db 363 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCFLGEECCYYVNSGIVTEKYKE 422
QY 421 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLNFVSSRI 480
Db 423 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLNFVSSRI 482
QY 481 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 538
Db 483 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 540

RESULT 4
ABP40988
ID ABP40988 standard; protein; 685 AA.
XX AC ABP40988;
XX DT 01-AUG-2002 (first entry)
XX DE Human retroviral env ORF protein.
XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
XX KW multiple sclerosis.
XX OS Human endogenous retrovirus.
XX PN WO9967395-A1.
XX PD 29-DEC-1999.
XX PF 23-JUN-1999; 99WO-FR001513.
XX PR 23-JUN-1998; 98FR-00007920.
XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Alliel PM, Perin J, Rieger F;
XX DR WPI; 2000-160587/14.
XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
XX PT for diagnosis, treatment and prevention of autoimmune and neurological
XX PT diseases.
XX PS Claim 22; Page 152-154; 225pp; French.
XX CC The present invention relates to new nucleic acid sequences of human
XX CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX CC Regulatory elements associated with HERV-7q may alter expression of other
XX CC genes (even remote genes) on the same chromosome, inducing immunological
XX CC and/or neurological changes (which may be pathological or protective/
XX CC curative). HERV-7q peptides can be used to improve efficiency of the
XX CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
XX CC sequences can be used in immunogenic or vaccinating compositions, for
XX CC protection against autoimmune diseases, particularly multiple sclerosis.
XX CC The peptides may also be used (by sequence comparison) to detect/identify
XX CC endogenous retroviruses that are abnormally expressed in cancer,
XX CC neuropathologies or other autoimmune diseases. The present sequence was
XX CC used to illustrate the invention
SQ Sequence 685 AA;

Query Match 100.0%; Score 538; DB 3; Length 685;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSPYQEFILWQMQRGNIDAPYSLSKGTP 60
Db 131 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSPYQEFILWQMQRGNIDAPYSLSKGTP 190

QY 61 TPTATHMPRCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVTCWYFTQTGMSDGGV 120
Db 191 TPTATHMPRCYHSATLCMHANTHYWTGKMINPSCPGGLGVTVTCWYFTQTGMSDGGV 250
QY 121 QDAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLTHTLRTLRVLSLNTLTGLHEVSA 180
Db 251 QDAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLTHTLRTLRVLSLNTLTGLHEVSA 310
QY 181 QNPTNCWICLPLNFRPYVSIIPVEQMNPFSTEINTSVLVGPLVSNLSTHTSNTLCVKF 240
Db 311 QNPTNCWICLPLNFRPYVSIIPVEQMNPFSTEINTSVLVGPLVSNLSTHTSNTLCVKF 370
QY 241 SNTTITNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSMCFLSFLVPPMTIY 300
Db 371 SNTTITNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCLNGSSSMCFLSFLVPPMTIY 430
QY 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYFKLSQELNGDMR 360
Db 431 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYFKLSQELNGDMR 490
QY 361 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCFLGEECCYYVNSGIVTEKYKE 420
Db 491 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCFLGEECCYYVNSGIVTEKYKE 550
QY 421 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLNFVSSRI 480
Db 551 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLNFVSSRI 610
QY 481 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 538
Db 611 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 668

RESULT 5
ABP41050
ID ABP41050 standard; protein; 846 AA.
XX AC ABP41050;
XX DT 01-AUG-2002 (first entry)
XX DE HERV-7q env protein #1.
XX DE Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
XX KW multiple sclerosis.
XX OS Human endogenous retrovirus.
XX PN WO9967395-A1.
XX PD 29-DEC-1999.
XX PF 23-JUN-1999; 99WO-FR001513.
XX PR 23-JUN-1998; 98FR-00007920.
XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Alliel PM, Perin J, Rieger F;
XX DR WPI; 2000-160587/14.
XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
XX PT for diagnosis, treatment and prevention of autoimmune and neurological
XX PT diseases.
XX PS Example 4; Fig 19; 225pp; French.
XX CC The present invention relates to new nucleic acid sequences of human
XX CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX CC Regulatory elements associated with HERV-7q may alter expression of other
XX CC genes (even remote genes) on the same chromosome, inducing immunological
XX CC and/or neurological changes (which may be pathological or protective/
XX CC curative). HERV-7q peptides can be used to improve efficiency of the
XX CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
XX CC sequences can be used in immunogenic or vaccinating compositions, for
XX CC protection against autoimmune diseases, particularly multiple sclerosis.
XX CC The peptides may also be used (by sequence comparison) to detect/identify
XX CC endogenous retroviruses that are abnormally expressed in cancer,
XX CC neuropathologies or other autoimmune diseases. The present sequence was
XX CC used to illustrate the invention
SQ Sequence 846 AA;

CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention
XX
XX
SQ Sequence 846 AA;

Query Match 100.0%; Score 538; DB 3; Length 846;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHFLFTVLPSFTLTAPPCCMTSSSPYQBFELMRMORPGNIDAPSVRSLSKGRP 60
DB 291 MALPYHFLFTVLPSFTLTAPPCCMTSSSPYQBFELMRMORPGNIDAPSVRSLSKGRP 350
QY 61 TPTAHTMPRNCYHSATLCHMANTHYWTGMINPSPCGGLGVTVCTYFTQGMSDGGV 120
DB 351 TPTAHTMPRNCYHSATLCHMANTHYWTGMINPSPCGGLGVTVCTYFTQGMSDGGV 410
QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRLTHRLVSLFNTLTGLHEVSA 180
DB 411 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRLTHRLVSLFNTLTGLHEVSA 470
QY 181 QNPTNCWICLPLNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEITHTSNLTCKVF 240
DB 471 QNPTNCWICLPLNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEITHTSNLTCKVF 530
QY 241 SNTTYTNSQCIRWVTPTQIVCLPSGIFPVCGTSAYRCLNGSSESMCFSLFVPPMTIY 300
DB 531 SNTTYTNSQCIRWVTPTQIVCLPSGIFPVCGTSAYRCLNGSSESMCFSLFVPPMTIY 590
QY 301 TEQDLYSYVISKPRNKRPVILPVIAGVLGALGTGIGITTSQFYKLSQELNGDME 360
DB 591 TEQDLYSYVISKPRNKRPVILPVIAGVLGALGTGIGITTSQFYKLSQELNGDME 650
QY 361 VADSLVTLQDQLNSLAVALVQNRALDLTAERGTCFLGEECCYVYVNSQGIIVTEKYKE 420
DB 651 VADSLVTLQDQLNSLAVALVQNRALDLTAERGTCFLGEECCYVYVNSQGIIVTEKYKE 710
QY 421 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGLAAIILLLFGPCIFNLLVNFVSSRI 480
DB 711 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGLAAIILLLFGPCIFNLLVNFVSSRI 770
QY 481 EAVKLQWEPKQSKTIYRPLDRPASPSDVNDIKGTPPEEISAAQPLLRPNAGSS 538
DB 771 EAVKLQWEPKQSKTIYRPLDRPASPSDVNDIKGTPPEEISAAQPLLRPNAGSS 828

RESULT 6

AA08622
ID AA08622 standard; protein; 538 AA.
XX
AC AA08622;
XX
DT 10-AUG-1999 (first entry)
XX
DE Human secreted protein A1172_2.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokine;
KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth.
XX
OS Homo sapiens.
XX
PN WO926972-A1.
XX

PD 03-JUN-1999.
XX
XX 17-NOV-1998; 98MO-US024614.
XX
PR 21-NOV-1997; 97US-00376110.
PR 18-MAY-1998; 98US-00080478.
PR 20-OCT-1998; 98US-00175928.
XX
XX (GEM) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M;
XX WPI; 1999-357813/30.
DR N-PSDB; AAX77526.
XX
XX New polynucleotides encoding secreted proteins.
PS Claim 14a; Page 101-103; 142pp; English.
XX
XX This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful for
CC gene therapy
XX
XX Sequence 538 AA;

Query Match 81.2%; Score 437; DB 2; Length 538;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MALPYHFLFTVLPSFTLTAPPCCMTSSSPYQBFELMRMORPGNIDAPSVRSLSKGRP 60
DB 1 MALPYHFLFTVLPSFTLTAPPCCMTSSSPYQBFELMRMORPGNIDAPSVRSLSKGRP 60
QY 61 TPTAHTMPRNCYHSATLCHMANTHYWTGMINPSPCGGLGVTVCTYFTQGMSDGGV 120
DB 61 TPTAHTMPRNCYHSATLCHMANTHYWTGMINPSPCGGLGVTVCTYFTQGMSDGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRLTHRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRLTHRLVSLFNTLTGLHEVSA 180
QY 181 QNPTNCWICLPLNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEITHTSNLTCKVF 240
DB 181 QNPTNCWICLPLNFRPVYSIPVPEQWNNFSTEINTSVLGPVLSNLEITHTSNLTCKVF 240
QY 241 SNTTYTNSQCIRWVTPTQIVCLPSGIFPVCGTSAYRCLNGSSESMCFSLFVPPMTIY 300
DB 241 SNTTYTNSQCIRWVTPTQIVCLPSGIFPVCGTSAYRCLNGSSESMCFSLFVPPMTIY 300
QY 301 TEQDLYSYVISKPRNKRPVILPVIAGVLGALGTGIGITTSQFYKLSQELNGDME 360
DB 301 TEQDLYSYVISKPRNKRPVILPVIAGVLGALGTGIGITTSQFYKLSQELNGDME 360
QY 361 VADSLVTLQDQLNSLAVALVQNRALDLTAERGTCFLGEECCYVYVNSQGIIVTEKYKE 420
DB 361 VADSLVTLQDQLNSLAVALVQNRALDLTAERGTCFLGEECCYVYVNSQGIIVTEKYKE 420
QY 421 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGLAAIILLLFGPCIFNLLVNFVSSRI 480
DB 421 IRDRIQRAEELRNTGPGWLLSQWMPWILPFLGLAAIILLLFGPCIFNLLVNFVSSRI 480

QY 481 EAVKLQMEPKQSKTKIYRRPLDRPAGPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538
 DB 481 EAVKLQMEPKQSKTKIYRRPLDRPAGPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538

RESULT 7
 ID AAY67313 standard; protein; 538 AA.
 AC AAY67313;
 DT 11-APR-2000 (first entry)
 DE Human secreted protein AJ172_2 amino acid sequence.
 KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
 KW placental pathology; metastasis inhibition; nutritional activity;
 KW immune stimulator; haematopoiesis regulator; tissue growth;
 KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115.
 OS Homo sapiens.
 PN WO9960020-A1.
 PD 25-NOV-1999.
 PF 17-MAY-1999; 99WO-US010915.
 PR 18-MAY-1998; 98US-00080478.
 PR 20-OCT-1998; 98US-00175928.
 PA (GEMV) GENETICS INST INC.
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Mi S, Treacy M;
 DR WPI; 2000-116311/10.
 DR N-PSDB; AAZ59468.
 PT New polynucleotides encoding secreted cDNA libraries, used to develop
 PT products for the diagnosis and treatment of neoplastic disease.
 PS Claim 15; Page 108-110; 149pp; English.
 SS This is the human secreted protein AJ172_2 amino acid sequence, the
 CC polynucleotide encoding AJ172_2 was obtained from a human adult testes
 CC cDNA library. The invention relates to secreted human and murine
 CC proteins. The polynucleotides and proteins are predicted to have
 CC biological activities which would make them suitable for treating,
 CC preventing or ameliorating medical conditions in humans and animals.
 CC Detection of the levels of the proteins can be used for the diagnosis of
 CC e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate
 CC the expression or function of the proteins may be used for treating a
 CC neoplastic disease and inhibiting metastasis. Other suggested activities
 CC include nutritional activity (e.g. in feeds), cytokine and cell
 CC proliferation/differentiation activity, immune stimulating (e.g. as
 CC vaccines) or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The
 CC polynucleotide sequences are also stated to be useful for gene therapy

Query Match 81.2%; Score 437; DB 3; Length 538;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMQRPNIDAPSYRLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMQRPNIDAPSYRLSKGTP 60

QY 61 TFTAHTHPRNCYHSATLCMHANTHYWTGKMINSPCGGLGVTVYCWYTFOTGMSDGGV 120
 DB 61 TFTAHTHPRNCYHSATLCMHANTHYWTGKMINSPCGGLGVTVYCWYTFOTGMSDGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKHLHETLRLTHTRLVSLFNLTTLGLHEVSA 180
 DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKHLHETLRLTHTRLVSLFNLTTLGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPYYSIPVPEQWNNFSTEINTTSVLVGLPYSLNLEITHSTNLTCKVF 240
 DB 181 QNPTNCWICLPLNFRPYYSIPVPEQWNNFSTEINTTSVLVGLPYSLNLEITHSTNLTCKVF 240
 QY 241 SNTVTYTNISQIRWVTPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFSLVPPMTIY 300
 DB 241 SNTVTYTNISQIRWVTPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFSLVPPMTIY 300
 QY 301 TQDLYSVVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDWER 360
 DB 301 TQDLYSVVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDWER 360
 QY 361 VADSLVTLDQQLNSLAAVVLQNRRLDLTLTAERGTCCLFLGEECCYYVNOQSGIVTEKYKE 420
 DB 361 VADSLVTLDQQLNSLAAVVLQNRRLDLTLTAERGTCCLFLGEECCYYVNOQSGIVTEKYKE 420
 QY 421 IRDRIQRRAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSRI 480
 DB 421 IRDRIQRRAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSRI 480
 QY 481 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538
 DB 481 EAVKLQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538

RESULT 8
 ID AAE14540 standard; protein; 538 AA.
 AC AAE14540;
 DT 07-MAY-2002 (first entry)
 DE Human syncytin protein.
 KW Human; syncytin; pre-eclampsia; gestational trophoblast disorder;
 KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;
 KW envelope gene; human endogenous defective retrovirus; HERV-W.
 OS Homo sapiens.
 PN WO200204678-A2.
 PD 17-JAN-2002.
 PF 09-JUL-2001; 2001WO-US021719.
 PF 07-JUL-2000; 2000US-0216657P.
 PR (GEMV) GENETICS INST INC.
 PA Keith JC, McCoy JM, Mi S;
 PI WPI; 2002-171727/22.
 DR N-PSDB; AAD24195.
 DR XX
 PT Identifying a compound for treating a subject with or at risk of
 PT developing pre-eclampsia, comprises determining whether the expression or
 PT activity of syncytin in the cell is modulated in the presence of a test
 PT compound.
 PS Disclosure; Page 42-43; 43pp; English.
 CC The invention relates to identifying compounds which are modulators of

CC syncytin expression. The syncytin modulators are useful in diagnosis and
 CC treatment of pre-eclampsia and gestational trophoblast disorders (e.g.
 CC choriocarcinoma, hydatiform mole, placental site tumour and missed/
 CC incomplete abortion). Syncytin is a human gene derived from the envelope
 CC gene of human endogenous defective retrovirus, HERV-W. The present
 CC invention is based partly on the discovery that syncytin expression is
 CC dramatically reduced in pre-eclampsia, and is also mis-localised to the
 CC apical syncytiotrophoblast membrane. The present sequence is human
 CC syncytin protein
 XX
 CC
 XX Sequence 538 AA;

Query Match 81.2%; Score 437; DB 5; Length 538;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMQRGNIDAPSYRLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWRMQRGNIDAPSYRLSKGTP 60
 QY 61 TTTAHTHPRNCHYHATLCMHANTHYWTGKINPSCPGGLGVTVCTYFTGTGSDGGV 120
 DB 61 TTTAHTHPRNCHYHATLCMHANTHYWTGKINPSCPGGLGVTVCTYFTGTGSDGGV 120
 QY 121 QDAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETRLTRVLSLFTNTLTGLHEVSA 180
 DB 121 QDAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETRLTRVLSLFTNTLTGLHEVSA 180
 QY 181 QNPTNCWICLPNFRPYVSPVPEQWNNFSTEINTSVLGPLVSNLEITTSNLTCKVF 240
 DB 181 QNPTNCWICLPNFRPYVSPVPEQWNNFSTEINTSVLGPLVSNLEITTSNLTCKVF 240
 QY 241 SNTTNTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFPLVPPMTIY 300
 DB 241 SNTTNTTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSMCFSLFPLVPPMTIY 300
 QY 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDME 360
 DB 301 TEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDME 360
 QY 361 VADSLVTLODQNSLAANVLQNRALDILTAERGTCFLGEECCYVYVQSGIVTEKYKE 420
 DB 361 VADSLVTLODQNSLAANVLQNRALDILTAERGTCFLGEECCYVYVQSGIVTEKYKE 420
 QY 421 IRDRIQRAEELRNTGPGWLLSQWMEWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 DB 421 IRDRIQRAEELRNTGPGWLLSQWMEWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLQMEPMQSKTKIYRPLDRPASPSVDNDIKGTPPEISAAQPLLPNSAGSS 538
 DB 481 EAVKLQMEPMQSKTKIYRPLDRPASPSVDNDIKGTPPEISAAQPLLPNSAGSS 538

RESULT 9

ADCS38777
 ID ADC38777 standard; protein; 538 AA.

XX AC ADC38777;
 XX AC ADC38777;

DT 18-DEC-2003 (first entry)

XX DE Human secreted protein #63.

XX immune disorder; severe combined immunodeficiency; SCID;
 KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
 KW rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
 KW lymphoid cell deficiency; osteoporosis; osteoarthritis;
 KW peripheral nervous system disease; peripheral neuropathy;
 KW Alzheimer's disease; Parkinson's disease; coagulation disorder;
 KW inflammatory disease; systemic inflammatory response syndrome; SIRS;
 KW ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
 KW hypersensitivity; regeneration; neural cell proliferation; fertility;
 KW tumour; chemokine; human; secreted protein.

XX Homo sapiens.
 XX US2002193567-A1.
 XX 19-DEC-2002.

XX 02-APR-2002; 2002US-00114893.

XX 11-AUG-1995; 95US-00514014.
 PR 05-APR-1996; 96US-00628364.
 PR 19-APR-1996; 96US-00635311.
 PR 07-JUN-1996; 96US-00659224.
 PR 17-JUN-1996; 96US-00664596.
 PR 09-JUL-1996; 96US-00677231.
 PR 26-JUL-1996; 96US-00686878.
 PR 23-AUG-1996; 96US-00701819.
 PR 27-SEP-1996; 96US-00721488.
 PR 27-SEP-1996; 96US-00721798.
 PR 27-SEP-1996; 96US-00721923.
 PR 27-SEP-1996; 96US-00721926.
 PR 25-OCT-1996; 96US-00738367.
 PR 30-OCT-1996; 96US-00739775.
 PR 13-JAN-1997; 97US-00783395.
 PR 10-APR-1997; 97US-00833823.
 PR 02-JUN-1997; 97US-00867677.
 PR 05-SEP-1997; 97US-00924838.
 PR 06-OCT-1999; 99US-00413232.

XX (GEMY) GENETICS INST INC.

XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Duckett M;
 PI Kelleher K;

DR WPI; 2003-657236/62.
 DR N-PSDB; ADC38776.

Proteins AZ3021 encoded by clone AZ3021 from human adult-colon, and
 BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
 useful for treating e.g. multiple sclerosis and rheumatoid arthritis.

Disclosure; SEQ ID NO 135; 412pp; English.

XX The invention relates to a protein comprising fully defined AZ302 1
 CC protein or BD127 1 6 protein. The polynucleotides are useful for
 CC expressing recombinant proteins for analysis and are also useful as
 CC chromosome markers or tags to identify chromosomes or to map related gene
 CC positions. The proteins are useful as amino acid supplement, carbon
 CC source, nitrogen source and carbohydrate source. The proteins are useful
 CC for treating various immune deficiencies and disorders (e.g. severe
 CC combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
 CC reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
 CC osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
 CC peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
 CC coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
 CC response syndrome (SIRS)), ischaemia-reperfusion injury, Crohn's disease,
 CC anaphylaxis and hypersensitivity. Proteins are also useful for inducing
 CC tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
 CC nerve growth or regeneration, for proliferating neural cells and for
 CC regenerating nerve and brain tissue, for inducing fertility and for
 CC inhibiting tumour growth. Proteins are also useful as chemokine for
 CC mammalian cells (e.g. monocytes, fibroblasts, neutrophils), and also
 CC useful as inhibitors of receptor/ligand interactions. The present
 CC sequence represents the amino acid sequence of a human secreted protein.

XX Sequence 538 AA;

Query Match 81.2%; Score 437; DB 7; Length 538;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 537; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSPYQEFLEWQRPNGNIDAPSYSLSGTIP 60
Db 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSPYQEFLEWQRPNGNIDAPSYSLSGTIP 60
QY 61 TFTAHTMPNCHYSAHLCHNANTHYTGKXINPSCGGGLGVTVCTWYFTQTGMSGGGV 120
Db 61 TFTAHTMPNCHYSAHLCHNANTHYTGKXINPSCGGGLGVTVCTWYFTQTGMSGGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLKSLHETLRTHTRVLSLENTTLGLHEVSA 180
Db 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLKSLHETLRTHTRVLSLENTTLGLHEVSA 180
QY 181 QNPNCWICPLNPRPYVSIPIVPOQNNFSTEINTSVLVGPLVSNLEIHTSNLTCVKF 240
Db 181 QNPNCWICPLNPRPYVSIPIVPOQNNFSTEINTSVLVGPLVSNLEIHTSNLTCVKF 240
QY 241 SNTTYTNSCIRWVTPTQVCLPSGIFVCGTSAVRCCLNGSSEMCFLSFLVPPMTIY 300
Db 241 SNTTYTNSCIRWVTPTQVCLPSGIFVCGTSAVRCCLNGSSEMCFLSFLVPPMTIY 300
QY 301 TEQDLYSVISKPRKRPVILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDME 360
Db 301 TEQDLYSVISKPRKRPVILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDME 360
QY 361 VADSLVTLQDLNLSAAVQLQNRDALDLTAERGCTCLFGEBCCCYYNOSGIVTEKYKE 420
Db 361 VADSLVTLQDLNLSAAVQLQNRDALDLTAERGCTCLFGEBCCCYYNOSGIVTEKYKE 420
QY 421 IRDRIQRAEBELRNTGPGWLLSQMWPWILPPLGFLAAIILLFPGCFIENLLVNFVSSRI 480
Db 421 IRDRIQRAEBELRNTGPGWLLSQMWPWILPPLGFLAAIILLFPGCFIENLLVNFVSSRI 480
QY 481 EAVKLQWEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 538
Db 481 EAVKLQWEPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLRPNSAGSS 538

RESULT 10

ADE09243
ID ADE09243 standard; protein; 544 AA.

AC ADE09243;
XX

DT 29-JAN-2004 (first entry)
XX

DE Novel protein-related contig polypeptide sequence #309.
XX

XX novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; contig.
XX

OS Unidentified.
XX

PN WO2003054152-A2.
XX

XX 03-JUL-2003.
XX

XX 10-DEC-2002; 2002WO-US039555.
XX

PR 10-DEC-2001; 2001US-0339733P.
PR

PR 11-DEC-2001; 2001US-0339453P.
PR

PR 14-MAR-2002; 2002US-0365091P.
PR

PR 14-MAR-2002; 2002US-0365384P.
PR

PR 12-APR-2002; 2002US-0372381P.
PR

PR 22-APR-2002; 2002US-0372615P.
PR

PR 24-APR-2002; 2002US-0376045P.
PR

PA (HYSE-) HYSEQ INC.
XX

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,
XX

DR WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.

XX Disclosure; SEQ ID NO 2787; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence was used in the
CC exemplification of the invention.

XX Sequence 544 AA;

Query Match 29.9%; Score 161; DB 7; Length 544;
Best Local Similarity 100.0%; Pred. No. 7.2e-144;
Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 LSKLHETLRTHTRVLSLENTTLGLHEVSAQNPTNCWICLPLNPRPYVSIPIVPOQNNFS 210

Db 156 LSKLHETLRTHTRVLSLENTTLGLHEVSAQNPTNCWICLPLNPRPYVSIPIVPOQNNFS 215

QY 211 TEINTSVLVGPLVSNLEIHTSNLTCVKFSNTTYTNSQCIRWVTPTQIVCLPSGIF 270

Db 216 TEINTSVLVGPLVSNLEIHTSNLTCVKFSNTTYTNSQCIRWVTPTQIVCLPSGIF 275

QY 271 VCGTSAVRCCLNGSSEMCFLSFLVPPMTIYTEQDLYSVIS 311

Db 276 VCGTSAVRCCLNGSSEMCFLSFLVPPMTIYTEQDLYSVIS 316

RESULT 11

AAM99345
ID AAM99345 standard; protein; 538 AA.

AC AAM99345;
XX

DT 21-MAY-1999 (first entry)
XX

DE Human endogenous retrovirus W. ORF1.
XX

XX Clone; human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker.

OS Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR001442.

XX 07-JUL-1997; 97FR-00008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W - expressed
PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
PT disease, and abnormal or failed pregnancy.
XX Example 8; Fig 6; 106pp; French.

XX This sequence represents the open reading frame 1 encoding the env

CC protein from the human endogenous retrovirus (HERV) W (AA025665). Nucleic
 CC acids, their fragments or peptides encoded by them derived from the HERV-
 CC W genome are markers of autoimmune disease (e.g. multiple sclerosis,
 CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
 CC dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility
 CC
 CC Sequence 538 AA;

Query Match 28.4%; Score 153; DB 2; Length 538;
 Best Local Similarity 100.0%; Pred. No. 2.8e-136;
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 EITHSNLTCVKSNTYITNSQIRWVTPPTQIVCLPSGIFVCGTSAAYRCLNGSSSM 287
 DB 228 EITHSNLTCVKSNTYITNSQIRWVTPPTQIVCLPSGIFVCGTSAAYRCLNGSSSM 287

QY 288 CFLSFLVPPWMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFY 347
 DB 288 CFLSFLVPPWMTIYTEQDLYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFY 347

QY 348 YKLSQELNGDMERVADSLVTLQDLNLSAAVVL 380
 DB 348 YKLSQELNGDMERVADSLVTLQDLNLSAAVVL 380

RESULT 12
 AAB75137
 ID AAB75137 standard; protein; 120 AA.
 AC AAB75137;
 XX
 XX 08-AUG-2001 (first entry)
 DT
 DE HERV-W truncated envelope protein GT.

XX Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
 KW envelope protein; multiple sclerosis-related superantigen; vaccine;
 KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
 KW antisense-therapy; autoimmune disorder.

XX Human endogenous retrovirus.
 OS
 XX WO200131021-A1.
 FN
 XX 03-MAY-2001.
 PD
 XX 30-OCT-2000; 2000WO-EP010659.
 PF
 XX 28-OCT-1999; 99EP-00402690.
 PR
 XX (UYGE-) UNIV GENEVE.
 XX
 XX Conrad B, Mach B;
 XX
 XX WPI; 2001-316336/33.
 DR

XX New human retrovirus HERV-W ENV proteins/peptides having superantigen
 PT activity useful for diagnosing and treating multiple sclerosis.
 PT
 XX
 XX Claim 1; Fig 7; 94pp; English.

XX On the basis of the PBS t-RNA motif used for the classification of human
 CC endogenous retrovirus (HERVs) the full length endogenous provirus which
 CC was located on the long arm of human chromosome 7 (7q21-22) has been
 CC designated HERV-W. The present invention describes proteins or peptides
 CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
 CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
 CC have neuroprotective activity, and can be used in: vaccines; antisense-
 CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
 CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated

CC disorders. (I) are also useful for identifying substances (and optionally
 CC recovering) capable of binding to a retroviral superantigen associated
 CC with MS, substances capable of blocking SAG activity and substances
 CC capable of blocking transcription or translation of HERV-W retroviral
 CC superantigen. A protein or peptide derived from (I), modified to be
 CC devoid of SAG activity and being capable of generating an immune response
 CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid
 CC molecules encoding (I) are useful as vaccines against MS. Substances
 CC capable of blocking SAG activity, capable of binding to a retroviral
 CC superantigen associated with MS, or capable of blocking transcription or
 CC translation of HERV-W retroviral superantigen for use in treating or
 CC preventing MS, obtained using (I) are useful for the treatment and
 CC prevention of MS. (I) and nucleic acids encoding them are useful for
 CC diagnosing autoimmune disease. The present sequence represents the
 CC specifically claimed truncated envelope protein of HERV-W designated GT
 XX
 XX Sequence 120 AA;

Query Match 22.3%; Score 120; DB 4; Length 120;
 Best Local Similarity 100.0%; Pred. No. 1.8e-105;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFELWQMORPGNIDAPSYLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFELWQMORPGNIDAPSYLSKGTP 60

QY 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKVINPSCPGGLGVTVCWTFYQTGMSDGGV 120
 DB 61 TFTAHTMPRNCYHSATLCMHANTHYWTGKVINPSCPGGLGVTVCWTFYQTGMSDGGV 120

RESULT 13
 ABG07532
 ID ABG07532 standard; protein; 446 AA.

XX AC ABG07532;
 XX
 XX 13-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #7523.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.
 OS
 XX WO200175067-A2.
 FN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 PR
 XX (HYSE-) HYSEQ INC.
 XX
 XX Drmanac RT, Liu C, Tang YT;
 XX
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS71719.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

XX Claim 20; SEQ ID NO 37891; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 446 AA;
 SQ Query Match 13.8%; Score 74; DB 4; Length 446;
 Best Local Similarity 100.0%; Pred. No. 2.9e-61;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 340 ITTSTQFYKLSOELNGDMERVADSLVTLOQLNSLAADVQLNRRALDLTAERGCTCLF 399
 Db 357 ITTSTQFYKLSOELNGDMERVADSLVTLOQLNSLAADVQLNRRALDLTAERGCTCLF 416
 QY 400 LGEECCYVNSGI 413
 Db 417 LGEECCYVNSGI 430

RESULT 14
 ABG20007
 ID ABG20007 standard; protein; 446 AA.
 XX AC ABG20007;
 XX DT 13-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #19998.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI; 2001-639362/73.
 XX DR N-PSDB; AAS84194.
 XX FT New isolated polynucleotide and encoded polypeptides, useful in
 XX FT diagnostics, forensics, gene mapping, identification of mutations
 XX FT responsible for genetic disorders or other traits and to assess
 XX FT biodiversity.
 XX PS Claim 20; SEQ ID NO 50366; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 446 AA;
 SQ Query Match 13.8%; Score 74; DB 4; Length 446;
 Best Local Similarity 100.0%; Pred. No. 2.9e-61;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 340 ITTSTQFYKLSOELNGDMERVADSLVTLOQLNSLAADVQLNRRALDLTAERGCTCLF 399
 Db 357 ITTSTQFYKLSOELNGDMERVADSLVTLOQLNSLAADVQLNRRALDLTAERGCTCLF 416
 QY 400 LGEECCYVNSGI 413
 Db 417 LGEECCYVNSGI 430

RESULT 15
 AAW42030
 ID AAW42030 standard; protein; 67 AA.
 XX AC AAW42030;
 XX DT 28-MAY-1998 (first entry)
 XX DE AJ172_2 protein.
 XX KW AJ172_2 clone; cDNA library; human adult testes; cell proliferation;
 XX KW cell differentiation.
 XX OS Homo sapiens.
 XX PN US5708157-A.
 XX PD 13-JAN-1998.
 XX PF 26-JUL-1996; 96US-00686878.
 XX PR 26-JUL-1996; 96US-00686878.
 XX PA (GEM) GENETICS INST INC.
 XX PI Mccoy JM, Treacy M, Racie LA, Merberg D, Spaulding V, Jacobs K;
 XX PI Lavallie ER, Evans C;
 XX DR WPI; 1998-109352/10.
 XX DR N-PSDB; AAV09124.
 XX FT Human testes derived cDNA and protein(s) - may be useful for laboratory
 XX FT use, or for stimulating cell proliferation and differentiation.
 XX PS Disclosure; Col 81-82; 48pp; English.
 XX CC This sequence is the AJ172-2 protein, which is 67 amino acids in length,
 CC and is isolated from a human adult testes cDNA library. When the AJ172_2

CC sequence was searched in the Genbank database, the clone demonstrated at
 CC least some homology with the following sequences: Friend murine leukemia
 CC virus; Moloney murine leukemia virus genome; pol protein, Gibbon leukemia
 CC virus; and an EST identified as "yh46a09.s1 Homo sapiens cDNA clone
 CC 132760 3,"

XX
 SQ Sequence 67 AA;

Query Match 12.5%; Score 67; DB 2; Length 67;
 Best Local Similarity 100.0%; Pred. No. 2.5e-55;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSTLTAPPCCMTSSSPYQEFLLWMQPCGNIDAPSVRSLSKGTG 60
 ID AAW44290 standard; protein; 67 AA.
 XX
 AC AAW44290;

QY 61 TPTATH 67
 ID AAW44290 standard; protein; 67 AA.

DB 61 TPTATH 67

RESULT 16
 AAW44290
 ID AAW44290 standard; protein; 67 AA.
 XX
 AC AAW44290;

XX
 DT 29-JUN-1998 (first entry)
 XX
 DE Human adult testes cDNA clone AJ172_2 protein sequence.

XX
 KW Human; adult; testes; AJ172_2; secreted protein; cytokine; nutrition;
 KW cell proliferation; differentiation; tumour; immune.
 XX
 OS Homo sapiens.

XX
 FN WO9804696-A1.
 XX
 PD 05-FEB-1998.

XX
 PF 23-JUL-1997; 97WO-US012910.

XX
 PR 26-JUL-1996; 96US-00686878.

XX
 PR 27-SEP-1996; 96US-00721489.

XX
 PA (GEM) GENETICS INST INC.

XX
 PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;
 PI Spaulding V;

XX
 DR WPI; 1998-130687/12.

XX
 DR N-PSDB; AAV15134.

XX
 PT Polynucleotides and secreted proteins having cytokine and cell
 PT proliferation/differentiation activity - isolated from a human adult
 PT testes cDNA library.

XX
 PS Claim 8; Page 34; 48pp; English.

XX
 CC The present sequence represents the protein sequence decoded from clone
 CC AJ172_2 from a human adult testes cDNA library. The nucleotide sequences
 CC and protein from clone AJ172_2, have uses in nutrition, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating or suppressing
 CC activity, haematopoiesis regulating activity, chemotactic/chemokinetic
 CC activity, haemostatic or thrombotic activity, receptor/ligand activity,
 CC anti-inflammatory activity, tumour inhibition activity or other
 CC activities. The products can also be used for production of antibodies
 CC and for detection and diagnosis

XX
 SQ Sequence 67 AA;

Query Match 12.5%; Score 67; DB 2; Length 67;
 Best Local Similarity 100.0%; Pred. No. 2.5e-55;

XX
 SQ Sequence 67 AA;

Query Match 12.5%; Score 67; DB 2; Length 67;
 Best Local Similarity 100.0%; Pred. No. 2.5e-55;

Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSTLTAPPCCMTSSSPYQEFLLWMQPCGNIDAPSVRSLSKGTG 60

DB 1 MALPYHIFLFTVLLPSTLTAPPCCMTSSSPYQEFLLWMQPCGNIDAPSVRSLSKGTG 60

QY 61 TPTATH 67

DB 61 TPTATH 67

RESULT 17

AAW99346

ID AAW99346 standard; protein; 52 AA.

XX
 AC AAW99346;

XX
 DT 21-MAY-1999 (first entry)

XX
 DE Human endogenous retrovirus W ORF2.

XX
 KW Clone; human endogenous retrovirus; genome; autoimmune disease;
 KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker.

XX
 OS Human endogenous retrovirus.

XX
 PN WO9902696-A1.

XX
 PD 21-JAN-1999.

XX
 PF 06-JUL-1998; 98WO-FR001442.

XX
 PR 07-JUL-1997; 97FR-00008815.

XX
 PA (INNR) BIO MERIEUX.

XX
 PI Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;

XX
 DR WPI; 1999-120897/10.

XX
 PT New nucleic acid sequences from human endogenous retrovirus-W - expressed
 PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
 PT disease, and abnormal or failed pregnancy.

XX
 PS Example 8; Fig 6; 106pp; French.

XX
 CC This sequence represents the open reading frame 2 from the human
 CC endogenous retrovirus (HERV) W (AA25665). Nucleic acids, their fragments
 CC or peptides encoded by them derived from the HERV-W genome are markers of
 CC autoimmune disease (e.g. multiple sclerosis, rheumatoid polyarthritis,
 CC disseminated lupus erythematosus, insulin-dependent diabetes and related
 CC pathologies) and of abnormal or unsuccessful pregnancy and can be used as
 CC chromosomal markers for susceptibility to these conditions, or proximity
 CC markers of genes associated with this susceptibility

XX
 SQ Sequence 52 AA;

Query Match 9.7%; Score 52; DB 2; Length 52;
 Best Local Similarity 100.0%; Pred. No. 3.6e-41;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 MEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 538

DB 1 MEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 52

RESULT 18

AAW71068

ID AAW71068 standard; protein; 493 AA.

XX
 AC AAW71068;

XX
 SQ Sequence 52 AA;

Query Match 9.7%; Score 52; DB 2; Length 52;
 Best Local Similarity 100.0%; Pred. No. 3.6e-41;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 MEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 538

DB 1 MEPMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 52

RESULT 18

AAW71068

ID AAW71068 standard; protein; 493 AA.

XX
 AC AAW71068;

XX
 SQ Sequence 52 AA;

DT 29-DEC-1998 (first entry)
 XX Multiple sclerosis associated retrovirus protein 3.
 XX
 XX
 XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
 KW env gene; rheumatoid arthritis-associated virus.
 XX
 XX Multiple sclerosis associated retrovirus.
 XX
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 39 /note= "Encoded by tga"
 FT
 XX
 XX WO9823755-A1
 XX
 XX 04-JUN-1998.
 XX
 XX 26-NOV-1997; 97WO-IB001482.
 XX
 XX 26-NOV-1996; 96US-00756429.
 PR
 XX (INMR) BIO MERIEUX.
 XX
 XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
 PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;
 XX
 XX WPI; 1998-322732/28.
 DR N-PSDB; AAV43217.
 XX
 XX New nucleic acid from retroviruses - useful for diagnosis, prevention and
 FT treatment of, e.g. multiple sclerosis.
 PT
 XX
 XX Disclosure; Page 185-187; 286pp; English.
 XX
 XX The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) protein used in the method of the invention. The
 CC invention provides complete or partial genomic sequences of the MSRV-1
 CC pol gene, gag gene and env gene, and polypeptides encoded by these genes.
 CC The invention also provides antibodies raised against the polypeptides.
 CC The genomic sequences, polypeptides and antibodies are also claimed
 CC useful for diagnosing infection by MS and rheumatoid arthritis-associated
 CC viruses, and also for prevention and treatment of infection with these
 CC viruses.
 CC
 XX Sequence 493 AA;
 SQ
 Query Match 8.4%; Score 45; DB 2; Length 493;
 Best Local Similarity 100.0%; Pred. No. 1.1e-33;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVFSNT 243
 DB |||||
 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVFSNT 243
 |||||
 RESULT 19
 AAW99553
 ID AAW99553 standard; protein; 493 AA.
 AC AAW99553;
 XX
 XX 17-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 08-JUN-1999 (first entry)
 XX
 XX Protein encoded by clone C15 from MSRV-1.
 DE
 XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KW rheumatoid polyarthritis.
 XX
 XX Viruses.
 OS
 XX
 XX Key Location/Qualifiers
 FT

FT Misc-difference 39 /label= unknown
 FT
 XX
 PN FR2765588-A1.
 XX
 PD 08-JAN-1999.
 XX
 XX 07-JUL-1997; 97FR-00008816.
 PF
 XX 07-JUL-1997; 97FR-00008816.
 PR
 XX (INMR) BIO MERIEUX.
 PA
 XX WPI; 1999-098275/09.
 DR N-PSDB; AAX29703.
 XX
 XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 PT multiple sclerosis or rheumatoid polyarthritis.
 XX
 XX Claim 2; Page 38-39; 83pp; French.
 PS
 XX This sequence represents the protein encoded by clone C15 from a novel
 CC multiple sclerosis related virus type 1 (MSRV1). The sequence can be used
 CC in diagnostic, prophylactic or therapeutic compositions to inhibit
 CC expression of a multiple sclerosis related virus and/or virus associated
 CC with rheumatoid polyarthritis. (Updated on 27-AUG-2003 to correct OS
 CC field.) (Updated on 17-OCT-2003 to standardise OS field)
 CC
 XX Sequence 493 AA;
 SQ
 Query Match 8.4%; Score 45; DB 2; Length 493;
 Best Local Similarity 100.0%; Pred. No. 1.1e-33;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVFSNT 243
 DB |||||
 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKVFSNT 243
 |||||
 RESULT 20
 AAB08195
 ID AAB08195 standard; protein; 541 AA.
 XX
 XX AAB08195;
 XX
 XX 06-AUG-2003 (revised)
 DT 04-DEC-2000 (first entry)
 DT
 XX
 XX Amino acid sequence of the MSRV-1 3' env and LTR regions.
 DE
 XX MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus.
 KW
 XX Multiple sclerosis associated retrovirus.
 OS
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..27 "signal peptide"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT Cleavage-site /note= "putative cleavage site"
 FT Cleavage-site /note= "putative cleavage site"
 FT Modified-site /note= "N-linked glycosylation site"
 FT

XX WO200047745-A1.
XX 17-AUG-2000.
XX
XX 15-FEB-2000; 2000WO-1B000159.
XX
XX 15-FEB-1999; 99EP-00420041.
XX (INMR) BIO MERIEUX.
XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX WPI; 2000-506097/45.
XX N-PSDB; AAA63826.
XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus
PT (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
PT sample.
XX
XX Disclosure; Fig 2; 23pp; English.
XX
XX The present sequence is encoded by a nucleotide sequence corresponding
CC to the 3' env region and long terminal repeat sequences from clone Cl6 of
CC Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
CC long terminal repeat (LTR)-RUS region which encodes the expression of a
CC MSRV-1 protein. This is unusual for LTRs, in particular in the RUS
CC region. Probes and antibodies to the MSRV-1 retrovirus protein and
CC encoding polynucleotide sequences are used to detect the presence of MSRV
CC -1 retrovirus in a biological sample. (Updated on 06-AUG-2003 to correct
CC OS field.)
XX
XX Sequence 541 AA;
XX
Query Match 8.4%; Score 45; DB 3; Length 541;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKFSNT 243
DB 198 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKFSNT 242
RESULT 21
AAB19069
ID AAB19069 standard; protein; 542 AA.
XX
XX AAB19069;
XX
XX 08-FEB-2001 (first entry)
XX
XX Amino acid sequence of an envelope (env) protein of MSRV-1.
XX
XX Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbeta16;
XX Vbeta17; multiple sclerosis; vaccine.
XX
XX Human endogenous retrovirus.
XX
XX WO200057185-A1.
XX
XX 28-SEP-2000.
XX
XX 20-MAR-2000; 2000WO-FR000691.
XX
XX 19-MAR-1999; 99FR-00003622.
XX 28-OCT-1999; 99FR-00013755.
XX (INMR) BIO MERIEUX.
XX
XX Perron H, Lafont M;
XX WPI; 2000-638279/61.
XX N-PSDB; AAA96625.
XX

XX Detecting superantigen activity, useful for identifying agents for
PT treatment or prevention of autoimmune disease, from expansion or loss of
PT particular lymphocyte Vbeta determinants.
XX
XX Claim 28; Page 124-126; 134pp; French.
XX
XX The present sequence represents an envelope protein of MSRV-1. The
CC envelope protein expressed by the endogenous human retrovirus MSRV-1 has
CC superantigen activity associated with autoimmune disease. The protein can
CC be detected using the method of the invention. The specification
CC describes a process for detecting activity of a superantigen in a
CC biological sample. The process comprises identifying large scale
CC expansion or loss of lymphocytes that carry at least one of the Vbeta16
CC and/or Vbeta17 determinants. The method is used to screen for agents that
CC inhibit the superantigen, especially those associated with MSRV-1 which
CC is implicated in autoimmune disease, particularly multiple sclerosis.
CC These agents are potentially useful for treatment or prevention (e.g. as
CC vaccines) of autoimmune diseases
XX
XX Sequence 542 AA;
XX
Query Match 8.4%; Score 45; DB 3; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.2e-33;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKFSNT 243
DB 199 SIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTCKFSNT 243
RESULT 22
ADD26900
ID ADD26900 standard; protein; 70 AA.
XX
XX ADD26900;
XX
XX 15-JAN-2004 (first entry)
XX
XX Human adipocyte Selected Interacting Domain (SID) prey protein #289.
XX
XX Adipocyte; protein-protein interaction; protein complex;
XX bait-prey complex; Selected Interacting Domain; SID; drug screening;
XX drug discovery; metabolic disease; obesity; lipodystrophy;
XX diabetes mellitus; type 2; non-insulin dependent; NIDDM;
XX adipogenesis modulation; gene therapy; human.
XX
XX Homo sapiens.
XX
XX WO200290544-A2.
XX
XX 14-NOV-2002.
XX
XX 03-MAY-2002; 2002WO-EP006333.
XX
XX 04-MAY-2001; 2001US-0288885P.
XX
XX (HYBR-) HYBRICENICS.
XX (LYNX-) LYNX THERAPEUTICS INC.
XX
XX Legrain P, Whiteside S, Mao J, Khrebukova I, Luo S;
XX WPI; 2003-111975/10.
XX N-PSDB; ADD26898.
XX
XX New complex of bait and prey between two polypeptides or polynucleotides
PT encoding the two polypeptides of adipocytes, useful for selecting a
PT modulating compound that inhibits or activates protein-protein
PT interactions.
XX
XX Claim 6; SEQ ID NO 289; 232pp; English.
XX
XX The invention relates to a bait-prey complex between two adipocyte
XX

CC polypeptides, or between two polynucleotides encoding adipocyte
CC polypeptides. The invention also relates to selected interacting domain
CC (SID) polypeptides which interact with selected bait polypeptides;
CC polynucleotides encoding SID polypeptides; vectors comprising SID
CC polynucleotides; recombinant host cells comprising an adipocyte
CC polynucleotide or a SID-encoding vector; a method of selecting for a
CC compound which modulates interactions between adipocyte polypeptides;
CC adipocyte modulator compounds identified using the method; a
CC pharmaceutical composition comprising an adipocyte modulator, or a SID-
CC encoding vector or host cell; and a protein chip comprising adipocyte
CC bait polypeptides. The bait-prey complexes of the invention are useful
CC for selecting a compound that inhibits or activates protein-protein
CC interactions between adipocyte polypeptides. The modulatory compounds
CC identified can be used in the treatment of metabolic diseases such as
CC obesity, lipodystrophy and type 2 diabetes mellitus, and in the
CC modulation of adipogenesis. The present sequence represents a
CC specifically claimed adipocyte SID prey polypeptide of the invention.

XX Sequence 70 AA;

Query Match 7.8%; Score 42; DB 7; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.5e-31;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 TSSSPYQEFLLWORMPGNIDAPSVRSLSKGTPTFTAHMTPR 70
|||||
DB 29 TSSSPYQEFLLWORMPGNIDAPSVRSLSKGTPTFTAHMTPR 70
|||||

RESULT 23

ABG08534
ID ABG08534 standard; protein; 144 AA.

XX AC ABG08534;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #8525.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS72721.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 38993; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: the sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

XX Sequence 144 AA;

Query Match 7.8%; Score 42; DB 4; Length 144;
Best Local Similarity 100.0%; Pred. No. 2.0e-31;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 SRIEAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTP 519
|||||
DB 84 SRIEAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTP 125
|||||

RESULT 24

ABG24204

ID ABG24204 standard; protein; 58 AA.

XX AC ABG24204;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #24195.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS88391.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 54563; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (ii) and its binding partners are useful in medical imaging
 CC of sites expressing (ii). (i) and (ii) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 58 AA;

Query Match 7.2%; Score 39; DB 4; Length 58;
 Best Local Similarity 100.0%; Pred. No. 8.9e-29;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 481 EAVKLQMEPKMOSKTKIYRPLDRPASPSVDNDIKGTP 519
 Db 1 EAVKLQMEPKMOSKTKIYRPLDRPASPSVDNDIKGTP 39

RESULT 25
 ADB84401
 ID ADB84401 standard; protein; 493 AA.
 AC ADB84401;
 XX
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE MSRV-1 associated protein/peptide sequence #10.
 XX
 KW multiple sclerosis; rheumatoid arthritis; gag; pol;
 KW reverse transcriptase; ribonuclease H.
 XX
 OS Unidentified;
 XX
 PN US2003039664-A1.
 XX
 PD 27-FEB-2003.
 XX
 PF 26-NOV-1997; 97US-00979847.
 XX
 PR 26-NOV-1996; 96US-00756429.
 XX
 PA (PERE/) PERRON H.
 PA (BESE/) BESEME F.
 PA (BEDI/) BEDIN F.
 PA (PARA/) PARANHOS-BACCALA G.
 PA (KOMU/) KOMURIAN-PRADEL F.
 PA (JOLI/) JOLIVET-REYNAUD C.
 PA (MAND/) MANDRAND B.
 PA (GARS/) GARSON J A.
 PA (TUKE/) TUKE P W.
 XX
 XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
 PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
 PI WPI; 2003-512253/43.
 DR
 XX
 XX New isolated or purified nucleic acid associated with multiple sclerosis
 PT and/or rheumatoid arthritis, useful for detecting a virus associated with
 PT multiple sclerosis or rheumatoid arthritis in a biological sample.
 XX
 XX Claim 31; Page 78-80; 193pp; English.
 PS
 XX The invention relates to an isolated or purified nucleic acid from a
 CC virus associated with multiple sclerosis and/or rheumatoid arthritis,
 CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
 CC pol, gag or reverse transcriptase genes (or their fragments) encoding the

CC proteins or defined peptides (including immunodominant peptides,
 CC antigenic peptides or conserved motifs). Also included are a process for
 CC detecting a virus associated with multiple sclerosis or rheumatoid
 CC arthritis in a biological sample, a nucleic acid probe for the detection
 CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a
 CC primer for the amplification by polymerisation of a nucleic acid of a
 CC viral material associated with multiple sclerosis or rheumatoid
 CC arthritis, a polypeptide exhibiting an inhibitory activity on the
 CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
 CC and an antibody directed against the MSRV-1 virus obtained by an
 CC immunologically reacting a human or animal body or cells with an
 CC immunogenic agent consisting of the antigenic polypeptide defined above.
 CC The nucleic acids are useful for detecting a biological sample a virus
 CC associated with multiple sclerosis or rheumatoid arthritis, or for
 CC detecting in a biological sample, the presence of or exposure to a virus
 CC associated with multiple sclerosis or rheumatoid arthritis. The present
 CC sequence is a claimed MSRV-associated sequence whose identity cannot be
 CC accurately determined. Note: The SEQ ID numbers for the sequences as
 CC displayed in the main body of the patent do not match the SEQ ID numbers
 CC in the sequence listing. Consequently those sequences mentioned in the
 CC claims may not be the sequences the authors intended to claim.
 XX
 SQ Sequence 493 AA;

Query Match 6.5%; Score 35; DB 7; Length 493;
 Best Local Similarity 100.0%; Pred. No. 3.6e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 209 FSTEINTTSLVGLVPLVSLNLEITHTSNTLCVKFSNT 243
 Db 209 FSTEINTTSLVGLVPLVSLNLEITHTSNTLCVKFSNT 243

RESULT 26
 ABP41044
 ID ABP41044 standard; peptide; 32 AA.
 AC ABP41044;
 XX
 DT 01-AUG-2002 (first entry)
 XX
 DE HERV-7q CKS-17 peptide.
 XX
 KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis.
 XX
 OS Human endogenous retrovirus.
 XX
 PN WO9967395-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 23-JUN-1999; 99WO-FR001513.
 XX
 PR 23-JUN-1998; 98FR-00007920.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Alliel PM, Perin J, Rieger F;
 XX WPI; 2000-160587/14.
 DR
 XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 PT for diagnosis, treatment and prevention of autoimmune and neurological
 PT diseases.
 XX
 PS Disclosure; Fig 5; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/

curative). HERV-7q peptides can be used to improve efficiency of the immune response, e.g. in immunotherapy. HERV-7q peptides and their coding sequences can be used in immunogenic or vaccinating compositions, for protection against autoimmune diseases, particularly multiple sclerosis. The peptides may also be used (by sequence comparison) to detect/identify endogenous retroviruses that are abnormally expressed in cancer, CC neuropathologies or other autoimmune diseases. The present sequence was CC used to illustrate the invention

XX SQ Sequence 32 AA;

Query Match 5.9%; Score 32; DB 3; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.4e-22;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 NSLAAVQLNRRALDLTAERGCTCLFLGEC 404
DB 1 NSLAAVQLNRRALDLTAERGCTCLFLGEC 32

RESULT 27
AAE25054
ID AAE25054 standard; protein; 542 AA.
XX AC AAE25054;
XX DT 30-OCT-2002 (first entry)
XX DE Human EMBRY-2 protein.
XX KW Human; embryogenesis associated protein; AIDS; reproductive disorder;
XX KW infertility; endometriosis; endometrial tumour; inflammatory disorder;
XX KW autoimmune disorder; acquired immune deficiency syndrome; transgenic;
XX KW ovarian tumour; contact dermatitis; placenta disorder; preclampsia;
XX KW EMBRY-2; allergy; gene therapy.
XX OS Homo sapiens.
XX PH 1.
XX FT Peptide
XX FT /label= Signal_peptide
XX FT /note= Mature EMBRY-2 protein"
XX FT 263..484
XX FT /note= "ENV polypeptide (coat polypeptide) domain"
XX FT 455..473
XX FT /note= "Transmembrane domain"
XX FN WO200248362-A2.
XX PD 20-JUN-2002.
XX PF 14-NOV-2001; 2001WO-US043956.
XX PR 15-NOV-2000; 2000US-0249407P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Ramkumar J, Arvizu C;
XX DR WPI; 2002-537629/57.
XX DR N-PSDB; AAD41225.
XX PT New polypeptides of human embryogenesis associated proteins for screening
XX PT modulators useful for treating or preventing disorders e.g.
XX PT endometriosis, infertility, allergy, preclampsia.
XX PS Claim 57; Page 94-95; 97pp; English.
XX CC The invention relates to human embryogenesis associated proteins (EMERY)
XX CC and nucleic acid molecules encoding such proteins. EMERY sequences are
XX CC useful for screening modulators useful for treating or preventing
XX CC disorders associated with abnormal expression of EMBRY. The disorders

treated include reproductive disorders such as infertility, CC endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory CC disorder such as acquired immune deficiency syndrome (AIDS), allergies, CC contact dermatitis; disorders of the placenta such as preclampsia, CC abruptio placentae etc. Sequences of the invention are also useful for CC analysing a proteome of a tissue or a cell type. EMBRY proteins are CC useful as immunogens for preparing antibodies. Polynucleotides of the CC invention are useful for creating knockin humanised animals or transgenic CC animals to model human diseases. They are also used in gene therapy. The CC present sequence is human EMBRY-2 protein

XX SQ Sequence 542 AA;

Query Match 5.9%; Score 32; DB 5; Length 542;
Best Local Similarity 100.0%; Pred. No. 2.8e-21;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 TTNSQIRWVTPPTQIVCLPSGIFVCGTSAY 277
DB 246 TTNSQIRWVTPPTQIVCLPSGIFVCGTSAY 277

RESULT 28
ABG28311
ID ABG28311 standard; protein; 368 AA.
XX AC ABG28311;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #28302.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS92498.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 58670; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 369 AA;
 SQ

Query Match 5.8%; Score 31; DB 4; Length 369;
 Best Local Similarity 100.0%; Pred. No. 1.8e-20;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 RGGTCLFLGECGCCYYNQSGIVTEKVEIRD 423
 |||||
 Db 203 RGGTCLFLGECGCCYYNQSGIVTEKVEIRD 233
 |||||

RESULT 29
 ABG28306
 ID ABG28306 standard; protein; 633 AA.
 XX
 AC ABG28306;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #28297.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS92493.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 58665; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 633 AA;
 SQ

Query Match 5.8%; Score 31; DB 4; Length 633;
 Best Local Similarity 100.0%; Pred. No. 2.8e-20;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 393 RGGTCLFLGECGCCYYNQSGIVTEKVEIRD 423
 |||||
 Db 484 RGGTCLFLGECGCCYYNQSGIVTEKVEIRD 514
 |||||

RESULT 30
 ABP41042
 ID ABP41042 standard; peptide; 29 AA.
 XX
 AC ABP41042;
 XX
 DT 01-AUG-2002 (first entry)
 XX
 DE Human retroviral CKH peptide.
 XX
 KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis.
 XX
 OS Human endogenous retrovirus.
 XX
 PN WO9967395-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 23-JUN-1999; 95WO-FR001513.
 XX
 PR 23-JUN-1998; 98FR-00007920.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Alliel PM, Perin J, Rieger F;
 XX
 DR WPI: 2000-160587/14.
 XX
 PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 PT for diagnosis, treatment and prevention of autoimmune and neurological
 PT diseases.
 XX
 PS Claim 26; Page 32; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention

XX Sequence 29 AA;
 SQ

Query Match 5.4%; Score 29; DB 3; Length 29;
 Best Local Similarity 100.0%; Pred. No. 1.6e-19;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Mon Apr 19 13:32:12 2004

Qy 380 LQNRALDLLTAERGTCFLGEECCYV 408
Db 1 LQNRALDLLTAERGTCFLGEECCYV 29

Search completed: April 19, 2004, 11:01:25
Job time : 62 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:58:55 ; Search time 21 Seconds
(without alignments)

2464.335 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 538

Sequence: 1 MALPWHPLFTVLLPSFTLTPPEISAAQPLLRNSAGSS 538

Scoring table:

OLIGO 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

PIR_78:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 9 | 1.7 | 570 | 4 B44282 | retrovirus-related |
| 2 | 8 | 1.5 | 45 | 2 PC1154 | tumor rejection an |
| 3 | 8 | 1.5 | 312 | 2 A80634 | probable 2-hydroxy |
| 4 | 8 | 1.5 | 328 | 2 E88930 | protein R11G11.11 |
| 5 | 8 | 1.5 | 669 | 2 C89904 | DNA mismatch repair |
| 6 | 8 | 1.5 | 1077 | 2 T41146 | probable cysteine- |
| 7 | 7 | 1.3 | 87 | 1 C6PSVM | cytochrome c5 - ps |
| 8 | 7 | 1.3 | 95 | 2 B75352 | conserved hypotet |
| 9 | 7 | 1.3 | 101 | 2 E75487 | v-type ATP synthas |
| 10 | 7 | 1.3 | 112 | 2 H2752 | hypothetical prote |
| 11 | 7 | 1.3 | 119 | 2 C71556 | holo-(acyl-carrier |
| 12 | 7 | 1.3 | 122 | 2 A86530 | acyl-carrier prote |
| 13 | 7 | 1.3 | 122 | 2 B72093 | acyl-carrier prote |
| 14 | 7 | 1.3 | 125 | 2 D81710 | holo-(acyl-carrier |
| 15 | 7 | 1.3 | 127 | 2 A84846 | hypothetical prote |
| 16 | 7 | 1.3 | 129 | 2 C40398 | phosphoribosylform |
| 17 | 7 | 1.3 | 133 | 2 D30445 | hypothetical prote |
| 18 | 7 | 1.3 | 137 | 2 S4755 | DNA-binding protei |
| 19 | 7 | 1.3 | 137 | 2 S00903 | DNA-binding protei |
| 20 | 7 | 1.3 | 137 | 2 C90846 | DNA-binding protei |
| 21 | 7 | 1.3 | 137 | 2 B85704 | DNA-binding protei |
| 22 | 7 | 1.3 | 140 | 2 T37059 | hypothetical prote |
| 23 | 7 | 1.3 | 144 | 1 S26725 | probable transcrip |
| 24 | 7 | 1.3 | 144 | 2 H82083 | conserved hypotet |
| 25 | 7 | 1.3 | 141 | 2 H95332 | protein [imported |
| 26 | 7 | 1.3 | 183 | 2 S73165 | H4-transporting tw |
| 27 | 7 | 1.3 | 183 | 2 I59442 | receptor tyrosine |
| 28 | 7 | 1.3 | 188 | 2 A98356 | hypothetical prote |
| 29 | 7 | 1.3 | 189 | 2 F89732 | protein ZK1073.2 (|

| | | | | |
|-----|-----|-----|----------|--------------------|
| 30 | 1.3 | 194 | 2 T27689 | hypothetical prote |
| 31 | 1.3 | 198 | 2 S60923 | hypothetical prote |
| 32 | 1.3 | 199 | 2 T39861 | mlo3 protein - fis |
| 33 | 1.3 | 201 | 2 A03985 | env polyprotein - |
| 34 | 1.3 | 204 | 2 C58930 | NADH2 dehydrogenas |
| 35 | 1.3 | 205 | 2 E72502 | hypothetical prote |
| 36 | 1.3 | 211 | 2 A03986 | env polyprotein - |
| 37 | 1.3 | 212 | 2 S13293 | KDEL receptor - hu |
| 38 | 1.3 | 212 | 2 A43994 | KDEL receptor - bo |
| 39 | 1.3 | 214 | 2 T51547 | hypothetical prote |
| 40 | 1.3 | 214 | 2 B84217 | hypothetical prote |
| 41 | 1.3 | 216 | 2 A11484 | transcription regu |
| 42 | 1.3 | 220 | 2 AH2212 | transport protein |
| 43 | 1.3 | 220 | 2 I80329 | receptor tyrosine |
| 44 | 1.3 | 227 | 2 C39925 | hypothetical prote |
| 45 | 1.3 | 229 | 2 AF2926 | hypothetical prote |
| 46 | 1.3 | 240 | 2 S59085 | ribosomal protein |
| 47 | 1.3 | 243 | 1 E64452 | conserved hypotet |
| 48 | 1.3 | 248 | 2 F69198 | conserved hypotet |
| 49 | 1.3 | 248 | 2 E71602 | probable integral |
| 50 | 1.3 | 249 | 1 QRECB | vitamin B12 transp |
| 51 | 1.3 | 249 | 2 D85779 | hypothetical prote |
| 52 | 1.3 | 249 | 2 A10704 | vitamin B12 ABC tr |
| 53 | 1.3 | 249 | 2 H90930 | hypothetical prote |
| 54 | 1.3 | 251 | 2 T15307 | hypothetical prote |
| 55 | 1.3 | 261 | 2 G87627 | hypothetical prote |
| 56 | 1.3 | 268 | 2 I53413 | calcium sensor pro |
| 57 | 1.3 | 272 | 1 S58290 | invasion-associate |
| 58 | 1.3 | 273 | 2 A42549 | hypothetical prote |
| 59 | 1.3 | 276 | 2 AD3486 | hypothetical prote |
| 60 | 1.3 | 280 | 2 S27835 | major surface anti |
| 61 | 1.3 | 282 | 2 T34070 | hypothetical prote |
| 62 | 1.3 | 283 | 2 JC4256 | hypothetical 32.0k |
| 63 | 1.3 | 286 | 2 F89796 | hypothetical prote |
| 64 | 1.3 | 287 | 2 D75177 | dolichyl-phosphate |
| 65 | 1.3 | 295 | 2 E81179 | spermidine/putresc |
| 66 | 1.3 | 295 | 2 A81927 | probable polyamine |
| 67 | 1.3 | 296 | 2 T23380 | hypothetical prote |
| 68 | 1.3 | 297 | 2 A45335 | cytochrome-c oxida |
| 69 | 1.3 | 303 | 2 E75106 | dihydroorotase deh |
| 70 | 1.3 | 305 | 2 S20600 | malM protein - Sal |
| 71 | 1.3 | 305 | 2 AG1014 | maltose operon per |
| 72 | 1.3 | 306 | 1 BVCEMM | maltose operon per |
| 73 | 1.3 | 306 | 2 D91256 | periplasmic protei |
| 74 | 1.3 | 306 | 2 H86096 | periplasmic protei |
| 75 | 1.3 | 306 | 2 S19997 | hypothetical prote |
| 76 | 1.3 | 308 | 2 C75125 | gmp synthase, cter |
| 77 | 1.3 | 313 | 1 A71037 | hypothetical prote |
| 78 | 1.3 | 317 | 2 F72246 | glucokinase - Ther |
| 79 | 1.3 | 321 | 2 S12812 | env polyprotein - |
| 80 | 1.3 | 322 | 2 S66766 | probable membrane |
| 81 | 1.3 | 325 | 2 A86202 | hypothetical prote |
| 82 | 1.3 | 332 | 2 G83066 | conserved hypotet |
| 83 | 1.3 | 335 | 2 T04029 | hypothetical prote |
| 84 | 1.3 | 335 | 2 A86464 | hypothetical prote |
| 85 | 1.3 | 338 | 2 D86815 | hypothetical prote |
| 86 | 1.3 | 342 | 2 T09355 | hypothetical prote |
| 87 | 1.3 | 342 | 2 B33061 | homeotic protein p |
| 88 | 1.3 | 347 | 2 B58888 | NADH2 dehydrogenas |
| 89 | 1.3 | 353 | 2 S22805 | env polyprotein - |
| 90 | 1.3 | 360 | 2 A84244 | hypothetical prote |
| 91 | 1.3 | 365 | 2 G83250 | chorismate mutase |
| 92 | 1.3 | 375 | 2 H70697 | probable DNA polym |
| 93 | 1.3 | 377 | 2 D72743 | probable DNA polym |
| 94 | 1.3 | 378 | 2 A00646 | spermidine/putresc |
| 95 | 1.3 | 382 | 2 T48109 | hypothetical prote |
| 96 | 1.3 | 387 | 2 T26735 | hypothetical prote |
| 97 | 1.3 | 388 | 2 AG3600 | hypothetical prote |
| 98 | 1.3 | 388 | 2 H97382 | probable transport |
| 99 | 1.3 | 389 | 2 T43709 | DNA topoisomerase |
| 100 | 1.3 | 389 | 2 T43710 | DNA topoisomerase |

ALIGNMENTS

RESULT 1

B44282
retrovirus-related env polyprotein pseudogene - human
N:Alternate names: coat polyprotein
N:Contains: coat protein gp70; coat protein p20E
C:Species: Homo sapiens (man)
C>Date: 17-Feb-1994 #sequence_revision 04-Jan-1996 #text_change 14-May-1999
C:Accession: B44282
R:Hirose, Y.; Takamatsu, M.; Harada, F.
Virology 192, 52-61, 1993

A:Title: Presence of env genes in members of the RTVL-H family of human endogenous retroviruses
A:Reference number: A44282, MUID:93297138, PMID:8517031
A:Accession: B44282

A:Status: conceptual translation of pseudogene
A:Molecule type: mRNA
A:Residues: 1-570 <HIR>

A:Cross-references: GB:D10083
C:Genetics:
A:Gene: env

C:Keywords: coat protein; polyprotein; pseudogene
F:1-372/Domain: coat protein gp70 <CP>
F:370/Region: opal stop codon
F:373-569/Domain: coat protein p20E <CPP>
F:436-448/Region: immunosuppressive peptide

Query Match 1.7%; Score 9; DB 4; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 374 SLAAVVLQN 382
| | | | |
DB 434 SLAAVVLQN 442

RESULT 2

PC1154
tumor rejection antigen P815 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 17-Mar-1999
C:Accession: PC1154
R:De Plaen, E.; Lurquin, C.; Van Pel, A.; Mariame, B.; Szikora, J.P.; Woelfel, T.; Sibil
Proc. Natl. Acad. Sci. U.S.A. 85, 2274-2278, 1988

A:Title: Immunogenic (tum-) variants of mouse tumor P815: Cloning of the gene of tum- an
A:Reference number: PC1154; MUID:88176927; PMID:3127830
A:Accession: PC1154
A:Molecule type: DNA
A:Residues: 1-45 <DEP>

A:Superfamily: catechol oxidase A2
Query Match 1.5%; Score 8; DB 2; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 381 QNRALDL 388
| | | | |
DB 15 QNRALDL 22

RESULT 3

AH0634
probable 2-hydroxyacid dehydrogenase in pho-csgg intergenic region STY1172 [imported]
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AH0634
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
Query Match 1.5%; Score 8; DB 2; Length 669;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AH0634

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-312 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08259.1; PID:g16502306; GSPDB:GN00176
C:Genetics:
A:Gene: STY1172
C:Superfamily: phosphoglycerate dehydrogenase

Query Match 1.5%; Score 8; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 IGAGVLGA 332
| | | | |
DB 142 IGAGVLGA 149

RESULT 4

E88930
Protein RL1G11.11 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: E88930
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_el
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an

A:Accession: E88930
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-328 <STO>
A:Cross-references: GB:chr_V; PIDN:AA69085.1; PID:g2384860; GSPDB:GN00023; CESP:RL1G11

C:Genetics:
A:Gene: RL1G11.11
A:Map position: 5

Query Match 1.5%; Score 8; DB 2; Length 328;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 FLFTVLLP 15
| | | | |
DB 2 FLFTVLLP 9

RESULT 5

C89904
DNA mismatch repair protein [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C89904
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogu
ma, A.; Mizutani-Tai, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hitamatsu, K.
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C89904
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-669 <KUR>
A:Cross-references: GB:BA000018; PID:g13701096; PIDN:BAB42391.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: mutL
C:Superfamily: mismatch repair protein hexB

Query Match 1.5%; Score 8; DB 2; Length 669;

```

Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 DVNDIKGT 518
    |||||
Db 462 DVNDIKGT 469

RESULT 6
T41146
probable cysteine-rich transcription regulator - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C/Accession: T41146
R/Hilbert, H.; Dueterhooft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A/Reference number: Z21973
A/Accession: T41146
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1077 <HIL>
A/Cross-references: EMBL:AL031907; PIDN:CAA21417.1; GSPDB:GN000068; SPDB:SPCC18.03
A/Experimental source: strain 972h-; cosmid c18
C/Genetics:
A/Gene: SPDB:SPCC18.03
A/Map position: 3
C/Superfamily: RING finger homology
F:193-252/Domain: RING finger homology <RRN>

Query Match 1.5%; Score 8; DB 2; Length 1077;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 KGLDLSKL 154
    |||||
Db 173 KGLDLSKL 180

RESULT 7
CCPSVM
cytochrome c5 - Pseudomonas mendocina
C/Species: Pseudomonas mendocina
C/Date: 07-May-1981 #sequence_revision 07-May-1981 #text_change 03-Mar-2000
C/Accession: A00114
R/Ambler, R.P.; Taylor, E.
Biochem. Soc. Trans. 1, 166-168, 1973
A/Title: Amino acid sequence of cytochrome c-5 from Pseudomonas mendocina.
A/Reference number: A00114
A/Accession: A00114
A/Molecule type: protein
A/Residues: 1-87 <AMB>
A/Experimental source: strain CH-110
C/Comment: This is a dimeric monoheme c-type cytochrome. It is unreactive with cytochrome
ic pseudomonads. It appears to be present also in the strictly anaerobic nitrogen-fixing
C/Superfamily: cytochrome c6; cytochrome c6 homology
C/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein
F:8-84/Domain: cytochrome c6 homology <CYC>
F:19/22/Binding site: heme (Cys) (covalent) #status predicted
F:23/63/Binding site: heme iron (His, Met) (axial ligands) #status predicted

Query Match 1.3%; Score 7; DB 1; Length 87;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 GAGVLGA 332
    |||||
Db 24 GAGVLGA 30

RESULT 8
B75352
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans

```

```

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C/Accession: B75352
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: B75352
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-95 <WHI>
A/Cross-references: GB:AE000513; GB:AE000513; NID:G6458395; PIDN:AAF10274.1; PID:G64584
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR1806
A/Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 95;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 IRDRIQR 427
    |||||
Db 84 IRDRIQR 90

RESULT 9
E75487
v-type ATP synthase, K subunit - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: E75487
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: E75487
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-101 <WHI>
A/Cross-references: GB:AE001926; GB:AE000513; NID:G6458395; PIDN:AAF10274.1; PID:G64584
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0696
A/Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 330 LGALGTG 336
    |||||
Db 49 LGALGTG 55

RESULT 10
H82752
hypothetical protein XF0867 [imported] - Xylella fastidiosa (strain 9a5c)
C/Species: Xylella fastidiosa
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C/Accession: H82752
R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A/Reference number: A82515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: H82752
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-112 <SIM>

```

A;Cross-references: GB:AE003926; GB:AE003849; NID:G9105771; PIDN:AAF83677.1; GSPDB:GN001
A;Experimental source: strain 945C
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Iwashiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
A;Genetics:
A;Gene: XF0867

Query Match 1.3%; Score 7; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 42; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0

QY 529 LLRPNSA 535
Db 19 LLRPNSA 25
|||||

RESULT 11
C71556
holo-[acyl-carrier-protein] synthase (EC 2.7.8.7) - Chlamydia trachomatis (serotype D, s
C;Species: Chlamydia trachomatis
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 05-May-2000
R;Accession: C71556
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A;Reference number: A71570; MUID:9900809; PMID:9784136
A;Accession: C71556
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-119 <ARN>
A;Cross-references: GB:AE001284; GB:AE001273; NID:G3328494; PIDN:AAC67691.1; PID:G332849
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: acps
C;Superfamily: holo-ACP synthase
C;Keywords: coenzyme A; transferase

Query Match 1.3%; Score 7; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 44; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0

QY 332 ALGTGIG 338
Db 61 ALGTGIG 57
|||||

RESULT 12
A86530
acyl-carrier protein synthase [imported] - Chlamydia pneumoniae (strain J138)
C;Species: Chlamydia pneumoniae
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
R;Accession: A86530
R;Shirai, M.; Harakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
Nucleic Acids Res. 28, 2311-2314, 2000
A;Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A;Reference number: A86491; MUID:20330349; PMID:10871362
A;Accession: A86530
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-122 <STO>
A;Cross-references: GB:BA000008; NID:G8978687; PIDN:BAA98523.1; GSPDB:GN00142

A;Experimental source: strain J138
C;Genetics:
A;Gene: acps
C;Superfamily: holo-ACP synthase

Query Match 1.3%; Score 7; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0

QY 332 ALGTGIG 338
Db 63 ALGTGIG 69
|||||

RESULT 13
B72093
acyl-carrier protein synthase - Chlamydia pneumoniae (strain CWL029)
C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
R;Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A;Reference number: A72000; MUID:9920606; PMID:10192388
A;Accession: B72093
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-122 <ARN>
A;Cross-references: GB:AE001616; GB:AE001363; NID:G4376584; PIDN:AAD18462.1; PID:G43765
A;Experimental source: strain CWL029
C;Genetics:
A;Gene: acps
C;Superfamily: holo-ACP synthase

Query Match 1.3%; Score 7; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 45; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0

QY 332 ALGTGIG 338
Db 63 ALGTGIG 69
|||||

RESULT 14
D81710
holo-(acyl-carrier protein) synthase TC0376 [imported] - Chlamydia muridarum (strain Ni
C;Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
R;Accession: D81710
R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A;Reference number: A81500; MUID:20150255; PMID:10684935
A;Accession: D81710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-125 <JET>
A;Cross-references: GB:AE002304; GB:AE002160; NID:G7190409; PIDN:AAF39234.1; PID:G71904
A;Experimental source: strain Nigg (MoPn)
C;Genetics:
A;Gene: TC0376
C;Superfamily: holo-ACP synthase

Query Match 1.3%; Score 7; DB 2; Length 125;
Best Local Similarity 100.0%; Pred. No. 46; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0

QY 332 ALGTGIG 338
Db 61 ALGTGIG 67
|||||

RESULT 15

A84846
 A:Accession: At2g41780 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 A:Accession: A84846
 R:Lin. X.; Kaul, S.; Rounselev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: A84846
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <STO>
 A:Cross-references: GB:AE002093; NID:G2335103; PIDN:AAC02773.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g41780
 A:Map position: 2

Query Match 1.3%; Score 7; DB 2; Length 127;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 418 VKEIRD 424
 |||||
 Db 117 VKEIRD 123

RESULT 16

C40398
 A:Accession: C40398
 A:Molecule type: DNA
 A:Residues: 1-129 <JEN>
 A:Cross-references: GB:IM59245; NID:G149726; PIDN:AAA72951.1; PID:G551892
 C:Genetics:
 A:Gene: purL
 C:Function:
 A:Description: catalyzes the condensation of 5'-phosphoribosylformylglycinamide with ATP
 A:Pathway: purine nucleotide biosynthesis
 C:Superfamily: phosphoribosylformylglycinamide synthase component II
 C:Keywords: ligase; purine nucleotide biosynthesis

Query Match 1.3%; Score 7; DB 2; Length 129;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GTGIGI 340
 |||||
 Db 93 GTGIGI 99

RESULT 17

D90445
 A:Accession: SSO2707 [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 A:Accession: D90445
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
 arett, R.A.; Ragan, M.A.; Sarsen, C.W.; Van der Oost, J.
 submitted to Genbank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139
 A:Accession: D90445
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-133 <KUR>
 A:Cross-references: GB:AE006641; NID:G13816025; PIDN:AAK42819.1; GSPDB:GN00155
 C:Genetics:
 A:Gene: SSO2707

Query Match 1.3%; Score 7; DB 2; Length 133;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 AITLLLL 463
 |||||
 Db 114 AITLLLL 120

RESULT 18

S24755
 A:Accession: S24755
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-137 <HRO>
 A:Cross-references: EMBL:X66848; NID:G47069; PIDN:CAA47322.1; PID:G47070
 A:Experimental source: strain 2a
 R:Yamada, M.; Sasaki, C.; Okada, N.; Makino, S.I.; Yoshikawa, M.
 Mol. Microbiol. 3, 207-213, 1989
 A:Title: Molecular cloning and characterization of Shigella virulence region kcpA of
 A:Reference number: S04173; MUID:89343624; PMID:2668687
 A:Accession: S04173
 A:Molecule type: DNA
 A:Residues: 1-109 <YAM>
 A:Cross-references: EMBL:X13131
 A:Experimental source: strain 2a; isolate YSH6200
 C:Genetics:
 A:Gene: virR; kcpA
 A:Map position: 27 min
 C:Function:
 A:Description: binds double-stranded DNA; required for coordinate temperature-regulated
 C:Superfamily: DNA-binding protein H-NS
 C:Keywords: DNA binding; transcription regulation

Query Match 1.3%; Score 7; DB 2; Length 137;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 LNSLAIV 378
 |||||
 Db 76 LNSLAIV 82

RESULT 19

S00903
 A:Accession: S00903
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-137 <HRO>
 A:Cross-references: EMBL:X66848; NID:G47069; PIDN:CAA47322.1; PID:G47070
 A:Experimental source: strain 2a
 R:Yamada, M.; Sasaki, C.; Okada, N.; Makino, S.I.; Yoshikawa, M.
 Mol. Microbiol. 3, 207-213, 1989
 A:Title: Molecular cloning and characterization of Shigella virulence region kcpA of
 A:Reference number: S04173; MUID:89343624; PMID:2668687
 A:Accession: S04173
 A:Molecule type: DNA
 A:Residues: 1-109 <YAM>
 A:Cross-references: EMBL:X13131
 A:Experimental source: strain 2a; isolate YSH6200
 C:Genetics:
 A:Gene: virR; kcpA
 A:Map position: 27 min
 C:Function:
 A:Description: binds double-stranded DNA; required for coordinate temperature-regulated
 C:Superfamily: DNA-binding protein H-NS
 C:Keywords: DNA binding; transcription regulation

Query Match 1.3%; Score 7; DB 2; Length 137;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 LNSLAIV 378
 |||||
 Db 76 LNSLAIV 82

A;Residues: 1-137 <PON>
A;Cross-references: EMBL:X07688; NID:g41735; PIDN:CAA30530.1; PID:g41736
A;Experimental source: strain MR5600
R;Goeransson, M.; Sonden, B.; Nilsson, P.; Dagberg, B.; Forsman, K.; Emanuelsson, K.; Uh
Nature 344, 682-685, 1990
A;Title: Transcriptional silencing and thermoregulation of gene expression in *Escherichia*
A;Reference number: S09325; MUID:90220876; PMID:1691451
A;Accession: S09325
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-137 <GOE>
R;Bremer, E.
submitted to the EMBL Data Library, December 1990
A;Reference number: S28633
A;Accession: S28633
A;Molecule type: DNA
A;Residues: 1-137 <BRE>
A;Cross-references: EMBL:X57231; NID:g42182; PIDN:CAA40507.1; PID:g42183
A;Experimental source: strain K-12
R;Palconi, M.; Gualtieri, M.T.; la Teana, A.; Losso, M.A.; Pon, C.L.
Mol. Microbiol. 2, 323-329, 1988
A;Title: Proteins from the prokaryotic nucleoid: primary and quaternary structure of the
A;Reference number: S03772; MUID:88288045; PMID:3135462
A;Accession: S03772
A;Molecule type: protein
A;Residues: 2-27, 'R', 29-57, 'I', 59-137 <PAL>
R;Ueguchi, C.; Ito, K.
J. Bacteriol. 174, 1454-1461, 1992
A;Title: Multicopy suppression: an approach to understanding intracellular functioning of
A;Reference number: A42290; MUID:92165719; PMID:1537791
A;Accession: S23789
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-137 <UEG>
A;Cross-references: EMBL:X59940; NID:g42025; PIDN:CAA42565.1; PID:g42027
R;Banchin, A.; Krin, E.
A;Experimental source: strain K-12
Microbiology 141, 959-960, 1995
A;Title: Filling the gap between hns and adhE in *Escherichia coli* K12.
A;Reference number: I57117; MUID:95291445; PMID:7773397
A;Accession: I57117
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-135 <RES>
A;Cross-references: EMBL:X67326; NID:g43077; PIDN:CAA47740.1; PID:g43078
A;Experimental source: strain K-12
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64870
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-137 <BLAT>
A;Cross-references: GB:AE000222; GB:U00096; NID:g1787486; PIDN:AACT4319.1; PID:g1787489;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: hns; drdx; oemZ; bgly
A;Map position: 27.5 min
C;Function:
A;Description: binds double-stranded DNA, with a preference for curved DNA; influences h
1 of pill production
A;Note: transcription enhanced through cold shock
C;Superfamily: DNA-binding protein H-NS
C;Keywords: DNA binding; nucleotide binding; P-loop; transcription regulation
F;2-137/Product: DNA-binding protein H-NS #status experimental <Mar>
F;122-129/Region: nucleotide-binding motif A (P-loop)

Query Match 1.3%; Score 7; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 LNSLAHV 378
|||||
Db 76 LNSLAHV 82

RESULT 20

C90846
DNA-binding protein H-NS [imported] - *Escherichia coli* (strain O157:H7, substrain RMD
C;Species: *Escherichia coli*
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C90846
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90846
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-137 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA035162.1; PID:g13361204; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: ECs1739
C;Superfamily: DNA-binding protein H-NS

Query Match 1.3%; Score 7; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 LNSLAHV 378
|||||
Db 76 LNSLAHV 82

RESULT 21

B85704
DNA-binding protein H-NS [similarity] - *Escherichia coli* (strain O157:H7, substrain EDL
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C;Accession: B85704
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85704
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-137 <STO>
A;Cross-references: GB:AE005174; NID:g12514958; PIDN:AAG56094.1; GSPDB:GN00145; UWGP:Z2
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: hns
C;Superfamily: DNA-binding protein H-NS

Query Match 1.3%; Score 7; DB 2; Length 137;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 372 LNSLAHV 378
|||||
Db 76 LNSLAHV 82

RESULT 22

T27059
Hypothetical protein Y51A2A.6 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27059
R;McMurray, A.
submitted to the EMBL Data Library, October 1998

A;Reference number: Z20304
A;Accession: T27059
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-140 <WIL>
A;Cross-references: EMBL:AL032635; PIDN:CAA21601.1; GSPDB:GN00023; CESP:Y51A2A.6
A;Experimental source: clone Y51A2A
C;Genetics:
A;Gene: CESP:Y51A2A.6
A;Map position: 5
A;Introns: 93/3; 129/1

Query Match 1.3%; Score 7; DB 2; Length 140;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 ILLLLFG 465
DB 7 ILLLLFG 13

RESULT 23
S26725
probable transcription termination factor nusA - Thermoplasma acidophilum
N;Alternate names: hypothetical protein X (tpoA2 3' region)
C;Species: Thermoplasma acidophilum
C;Date: 12-Feb-1993 #sequence_revision 10-Jul-1998 #text_change 16-Jul-1999
C;Accession: S26725
R;Klenk, H.P.; Renner, O.; Schwass, V.; Zillig, W.
Nucleic Acids Res. 20, 5226, 1992
A;Title: Nucleotide sequence of the genes encoding the subunits H, B, A' and A'' of the
A;Reference number: S26721; MUID:93027268; PMID:1408839
A;Accession: S26725
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-144 <KLE>
A;Cross-references: EMBL:X68198; NID:G48089; PIDN:CAA48283.1; PID:G48094
A;Experimental source: DSM 1728
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992
C;Superfamily: archaeobacterial probable transcription termination factor nusA
C;Keywords: transcription termination

Query Match 1.3%; Score 7; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 KEVISQL 136
DB 54 KEVISQL 60

RESULT 24
H82083
conserved hypothetical protein VC2378 [imported] - Vibrio cholerae (strain N16961 serog
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: H82083
R;Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
L; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: H82083
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-144 <HEI>
A;Cross-references: GB:AE004308; GB:AE003852; NID:G9656946; PIDN:AAF95521.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2378
A;Map position: 1

Query Match 1.3%; Score 7; DB 2; Length 144;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 RRAEELR 433
DB 126 RRAEELR 132

RESULT 25
H95332
protein [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSymA
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: H95332
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow
K; Kalman, S.; Keating, D.H.; Palm, C.; Peck, K.; Surzycki, R.; Wells, D.H.; Yeh, K.C
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilo
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: H95332
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-181 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK65226.1; PID:G14523675; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:
A;Gene: Sma1050
A;Genome: plasmid

Query Match 1.3%; Score 7; DB 2; Length 181;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 GVLGALG 334
DB 83 GVLGALG 89

RESULT 26
S73165
H+-transporting two-sector ATPase (EC 3.6.3.14) chain b - red alga (Porphyra purpurea)
C;Species: Chloroplast Porphyra purpurea
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
C;Accession: S73165
R;Reith, M.; Munnolland, J.
Plant Mol. Biol. Rep. 13, 333-335, 1995
A;Title: Complete nucleotide sequence of the Porphyra purpurea chloroplast genome.
A;Reference number: S73108
A;Accession: S73165
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-183 <REI>
A;Cross-references: EMBL:U38804; NID:G1276652; PIDN:AAC08130.1; PID:G1276710
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Gene: atpF
A;Genome: chloroplast
C;Superfamily: H+-transporting ATP synthase chain I
C;Keywords: Arp biosynthesis; chloroplast; hydrolase; membrane-associated complex; thyl.

Query Match 1.3%; Score 7; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 459 ILLLLFG 465
 |||||
 Db 37 ILLLLFG 43

RESULT 27

I59442
 receptor tyrosine kinase - rat (fragment)
 C/Species: Rattus sp. (rat)
 C/Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 24-Sep-1999
 C/Accession: J59442
 R/Sakuma S.; Sava H.; Iijichi, A.; Tofilon, P.J.
 Radiat. Res. 143, 1-7, 1995
 A/Title: Radiation induction of the receptor tyrosine kinase gene Ptk-3 in normal rat as
 A/Reference number: I59442; PMID:95320273; PMID:7597135
 A/Accession: I59442
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-183 <RES>
 A/Cross-references: GB:I577556; NID:G957334; PIDN:AA834729.1; PID:G957335
 C/Genetics:
 A/Gene: Ptk-38
 C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 1.3%; Score 7; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 AIIILL 463
 |||||
 Db 22 AIIILL 28

RESULT 28

A98356
 hypothetical protein AGR_L_3579 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C/Accession: A98356
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollan, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A/Reference number: A97359; PMID:21608551; PMID:11743194
 A/Accession: A98356
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-188 <KUR>
 A/Cross-references: GB:AE007870; PIDN:AAK90371.1; PID:G15160413; GSPDB:GN00170
 C/Genetics:
 A/Gene: AGR_L_3579
 A/Map position: linear chromosome

Query Match 1.3%; Score 7; DB 2; Length 188;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 FTVLLPS 16
 |||||
 Db 16 FTVLLPS 22

RESULT 29

F89732
 protein ZK1073.2 [imported] - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
 C/Accession: F89732
 R/anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
 A/Reference number: A75000; PMID:99069613; PMID:9851916

A/Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_el
 A/Note: Published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
 A/Accession: F89732
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-189 <STO>
 A/Cross-references: GB:chr_X; PIDN:CAA92228.1; PID:G3881512; GSPDB:GN00028; CESP:ZK1073
 C/Genetics:
 A/Gene: ZK1073.2
 A/Map position: X

Query Match 1.3%; Score 7; DB 2; Length 189;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 AAIIILL 462
 |||||
 Db 83 AAIIILL 89

RESULT 30

T27689
 hypothetical protein ZK1073.2 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C/Accession: T27689
 R/McMurray, A.
 submitted to the EMBL Data Library, November 1995
 A/Reference number: Z20405
 A/Accession: T27689
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-194 <WIL>
 A/Cross-references: EMBL:Z68135; PIDN:CAA92228.2; GSPDB:GN00028; CESP:ZK1073.2
 A/Experimental source: clone ZK1073
 C/Genetics:
 A/Gene: CESP:ZK1073.2
 A/Map position: X
 A/Introns: 4/2; 41/3; 111/3; 146/1

Query Match 1.3%; Score 7; DB 2; Length 194;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 AAIIILL 462
 |||||
 Db 88 AAIIILL 94

Search completed: April 19, 2004, 11:02:35
 Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:33:58 ; Search time 20 Seconds
(without alignments)
2587.552 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 2871

Sequence: 1 MALPPIHFLFTVLPSFTILT.....PPEISAAQPLLRNSAGSS 538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_78:**

1: PIR1:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 627 | 21.8 | 563 | 1 VCMVW7 | env polyprotein - |
| 2 | 599 | 20.9 | 587 | 1 VCLJSA | env polyprotein - |
| 3 | 593.5 | 20.7 | 567 | 1 VCFVAS | env polyprotein - |
| 4 | 593.5 | 20.7 | 586 | 1 VCLJMP | env polyprotein - |
| 5 | 549 | 19.1 | 575 | 1 VCLJHD | env polyprotein pr |
| 6 | 508.5 | 17.7 | 570 | 4 B44282 | retrovirus-related |
| 7 | 502 | 17.5 | 582 | 1 VCVDA8 | env polyprotein - |
| 8 | 465.5 | 16.2 | 321 | 2 S12812 | env polyprotein - |
| 9 | 424 | 14.8 | 640 | 1 VCMVRV | env polyprotein pr |
| 10 | 419 | 14.6 | 676 | 2 T01381 | env protein - muri |
| 11 | 418.5 | 14.6 | 565 | 1 VCVWEM | env polyprotein - |
| 12 | 414 | 14.4 | 676 | 2 S70395 | env polyprotein - |
| 13 | 413.5 | 14.4 | 688 | 2 A43491 | env polyprotein - |
| 14 | 413.5 | 14.4 | 689 | 2 B43491 | env polyprotein - |
| 15 | 412.5 | 14.4 | 662 | 1 VCVWGF | env polyprotein - |
| 16 | 412.5 | 14.4 | 662 | 2 A25982 | env polyprotein - |
| 17 | 411.5 | 14.3 | 662 | 1 VCMVLB | env polyprotein - |
| 18 | 410 | 14.3 | 676 | 1 VCMVPV | env polyprotein pr |
| 19 | 409 | 14.2 | 645 | 1 VCMVSS | env polyprotein pr |
| 20 | 408 | 14.2 | 668 | 1 VCMVFP | env polyprotein pr |
| 21 | 407.5 | 14.2 | 642 | 1 VCMVSP | env polyprotein pr |
| 22 | 406.5 | 14.2 | 636 | 1 VCMVFS | env polyprotein - |
| 23 | 406.5 | 14.2 | 642 | 2 T10533 | env polyprotein pr |
| 24 | 406.5 | 14.2 | 669 | 1 VCVWEX | env polyprotein - |
| 25 | 406 | 14.1 | 627 | 1 VCMVW2 | env polyprotein - |
| 26 | 403.5 | 14.1 | 642 | 1 VCMVFG | env polyprotein - |
| 27 | 402 | 14.0 | 640 | 1 VCVWM1 | env polyprotein - |
| 28 | 401.5 | 14.0 | 669 | 2 A46511 | envelope protein - |
| 29 | 401 | 14.0 | 665 | 1 VCMVVR | env polyprotein pr |

| | | | | | |
|-----|-------|------|------|-----------|-----------------------|
| 30 | 399.5 | 13.9 | 639 | 1 VCMVSA | env polyprotein pr |
| 31 | 399 | 13.9 | 665 | 1 VCMVKA | env polyprotein pr |
| 32 | 389.5 | 13.6 | 644 | 2 S15464 | gp70 protein - mur |
| 33 | 389.5 | 13.6 | 667 | 1 VCLJGL | env polyprotein pr |
| 34 | 387.5 | 13.5 | 661 | 1 VCMVCB | env polyprotein - |
| 35 | 386.5 | 13.5 | 353 | 2 S22805 | env polyprotein - |
| 36 | 384.5 | 13.4 | 666 | 1 VCMVHL | env polyprotein pr |
| 37 | 373 | 13.0 | 671 | 1 VCMVCE | env polyprotein - |
| 38 | 330.5 | 11.5 | 211 | 2 A03986 | env polyprotein - |
| 39 | 327.5 | 11.4 | 201 | 2 A03985 | env polyprotein - |
| 40 | 213.5 | 7.4 | 486 | 1 VCLJH2 | env polyprotein - |
| 41 | 213.5 | 7.4 | 486 | 1 VCLJH2 | envelope glycoprot |
| 42 | 209.5 | 7.3 | 488 | 2 A45714 | transmembrane enve |
| 43 | 209 | 7.3 | 247 | 2 A53482 | env polyprotein - |
| 44 | 206 | 7.2 | 488 | 1 VCLJCN | env polyprotein - |
| 45 | 203 | 7.1 | 488 | 1 VCLJMT | envelope glycoprot |
| 46 | 202 | 7.0 | 488 | 2 S14605 | env polyprotein - |
| 47 | 199 | 6.9 | 488 | 2 B61547 | env polyprotein - |
| 48 | 187.5 | 6.5 | 488 | 1 VCMVH | env polyprotein - |
| 49 | 187 | 6.5 | 173 | 2 B41281 | env polyprotein - |
| 50 | 187 | 6.5 | 173 | 2 A41281 | env polyprotein - |
| 51 | 186 | 6.5 | 173 | 2 F41281 | env polyprotein - |
| 52 | 184 | 6.4 | 173 | 2 B41281 | env polyprotein - |
| 53 | 181 | 6.3 | 173 | 2 C41281 | env polyprotein - |
| 54 | 176 | 6.1 | 534 | 1 VCVWSF | env polyprotein - |
| 55 | 174 | 6.1 | 173 | 2 H41281 | env polyprotein - |
| 56 | 167.5 | 5.8 | 515 | 1 VCLJGA | env polyprotein pr |
| 57 | 153.5 | 5.3 | 515 | 1 VCLJGB | env polyprotein - |
| 58 | 152 | 5.3 | 594 | 2 H48613 | env polyprotein pr |
| 59 | 152 | 5.3 | 606 | 2 S35427 | env polyprotein - |
| 60 | 148.5 | 5.2 | 603 | 1 VCFVER | env polyprotein - |
| 61 | 146 | 5.1 | 601 | 2 I48613 | env polyprotein pr |
| 62 | 141 | 4.9 | 681 | 2 A45705 | type I transmembra |
| 63 | 139 | 4.8 | 681 | 2 S33316 | structural protein |
| 64 | 130 | 4.5 | 257 | 2 A43362 | env polyprotein - |
| 65 | 126.5 | 4.4 | 246 | 1 VCFV37 | coat protein gp37 |
| 66 | 124 | 4.3 | 174 | 1 VCFVUR | env polyprotein TM |
| 67 | 122.5 | 4.3 | 198 | 2 C48613 | env polyprotein ev |
| 68 | 121 | 4.2 | 246 | 2 S23734 | env-sea polyprotei |
| 69 | 121 | 4.2 | 596 | 1 TVFVSA | env polyprotein, r |
| 70 | 120.5 | 4.2 | 432 | 2 A25483 | env polyprotein - |
| 71 | 117 | 4.1 | 246 | 2 A60398 | env polyprotein - sim |
| 72 | 116 | 4.0 | 77 | 2 A03982 | p15E protein - sim |
| 73 | 115.5 | 4.0 | 408 | 1 VCMVSR | env polyprotein - |
| 74 | 109.5 | 3.8 | 551 | 2 D64412 | hypothetical prote |
| 75 | 108.5 | 3.8 | 837 | 2 S29955 | surface glycoprote |
| 76 | 108.5 | 3.8 | 855 | 1 JQ2004 | env polyprotein - |
| 77 | 107 | 3.7 | 444 | 2 A46165 | envelope surface g |
| 78 | 106 | 3.7 | 856 | 2 S23823 | env polyprotein - |
| 79 | 105.5 | 3.7 | 318 | 2 S23155 | envelope glycoprot |
| 80 | 104.5 | 3.6 | 194 | 2 I50432 | transmembrane prot |
| 81 | 104.5 | 3.6 | 855 | 1 JQ2003 | env polyprotein pr |
| 82 | 104 | 3.6 | 854 | 1 VCLJSI | env polyprotein E |
| 83 | 104 | 3.6 | 857 | 2 S19886 | mucin 6, gastric (|
| 84 | 102 | 3.6 | 505 | 2 B46529 | env polyprotein pr |
| 85 | 101.5 | 3.5 | 990 | 1 G46335 | mucin MUC5B, trach |
| 86 | 101.5 | 3.5 | 3570 | 2 T45025 | envelope surface g |
| 87 | 101 | 3.5 | 408 | 2 C46165 | env polyprotein - |
| 88 | 101 | 3.5 | 445 | 1 VCFMLV | env polyprotein k |
| 89 | 101 | 3.5 | 1338 | 2 T18287 | protein-tyrosine k |
| 90 | 101 | 3.5 | 3005 | 2 S33642 | homeotic protein z |
| 91 | 100 | 3.5 | 403 | 2 B46165 | hypothetical prote |
| 92 | 100 | 3.5 | 1467 | 2 T23950 | env polyprotein pr |
| 93 | 99.5 | 3.5 | 966 | 1 VCLJCC | hypothetical prote |
| 94 | 99 | 3.4 | 421 | 2 T50138 | env polyprotein pr |
| 95 | 99 | 3.4 | 856 | 1 VCLJFP | gag polyprotein - |
| 96 | 99 | 3.4 | 863 | 2 A53034 | probable DNA repai |
| 97 | 99 | 3.4 | 897 | 2 T37813 | transcription co-r |
| 98 | 98.5 | 3.4 | 1455 | 2 S60255 | flagella-related p |
| 99 | 97.5 | 3.4 | 552 | 2 T44954 | env polyprotein pr |
| 100 | 97.5 | 3.4 | 964 | 1 VCLJOC6 | |

ALIGNMENTS

RESULT 1

VCWVW7

env polyprotein - baboon endogenous virus (strain M7)
N:Alternate names: coat polyprotein
N:Contains: coat protein gp70; coat protein p20E
C:Species: baboon endogenous virus
A:Note: host Papio sp. (baboon)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C:Accession: J02062
R:Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.
Jpn. J. Genet. 62, 127-137, 1987
A:Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome
F:437-449/Region: immunosuppressive peptide #status predicted
A:Reference number: J02060
A:Accession: J02062
A:Molecule type: DNA
A:Residues: 1-563 <XAT>
A:Cross-references: GB:M16550; NID:G509586; PIDN:AAA87333.1; PID:G332599
C:Genetics:

C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
F:1-376/Product: coat protein gp70 #status predicted <CGP>
F:377-563/Product: coat protein p20E #status predicted <CGP>
F:437-449/Region: immunosuppressive peptide #status predicted
F:113,219,229,264,282,292,306,312,321,339,469/Binding site: carbohydrate (Aen) (covalent

Query Match 21.8%; Score 627; DB 1; Length 563;
Best Local Similarity 31.6%; Pred. No. 5.8e-39;
Matches 158; Conservative 87; Mismatches 169; Indels 86; Gaps 16;

QY 31 SSFYQELFMQPGNIDAPSYSLKSGTPTFAHTHMPNCHSNTLCMHANTHWY--- 87
Db 83 SGPLOEC-----PCN-----SVHS-----SCYTSQQCRSGNKTYTAT 120
QY 88 -----TGMINPSPGGLGVTVCTYQTQGMDSGGGVQQAQAEKHVK 130
Db 121 LLXTQTGTSVQVLGSTNKLQSPGNGIKGQISCHSTTAPIHVSDDGGPLDTR---IX 177
QY 131 EVISQTRVHGTSSP---YKGLDLKSLHETLRTHTRLVSLFNLTGLHEVSAQPTNCW 187
Db 178 SVQRKLEIHKALYPELOXHLAIPKVRDNLMDVDAQTLLNATYNLLMSNTSLVDDCW 237
QY 188 ICLPLNRPVSPVPEQWNNPSTETI--NTTSVLGVPLVSNLEITHTSNLTG--- 237
Db 238 LCLKLG--PPTPLAIFNLSYTRSDNISCLIIIPLL--VQPMQFSNSSLFPSPYNS 293
QY 238 -----VKFSNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTS--AYRCLNGSSEMC 288
Db 294 TEBIDLGHVAFSCTSIITN-----VTGP-----ICAVNGSVFLCGNNMAYTLPNTWTGLC 344
QY 289 FLSFLVPPMTIYEQD-----LYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTS 343
Db 345 VLATLPLDIDIIPGDEVPPIADHFIYRPK-RAIQFIPLAGLTAATAFTGATGLGV 403
QY 344 TQFYKLSQELNGDMERVADSLVTLQDQLNSLAAYVLQNRRLDLLTAERGTCFLFLGEE 403
Db 404 VTQYTKLSNQLISDVQILSSTIQDQDQVDSLAEVVLQNRGLDILLTAEQGGICLALQEK 463
QY 404 CCYVQNSGIVTEKVEIRDRIORRAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLL 463
Db 464 CCFYVNSGIVRDKITLQEELEERRKDLASNLWTGLOGLLPYLLPFLPLTLLLT 523
QY 464 FGPICIFNLLVNFSSRIEAV 483
Db 524 IGPICIFNRLTAFINDKLINII 543

RESULT 2

VCLJSA

env polyprotein - simian AIDS retrovirus SRV-1

N:Alternate names: coat polyprotein
N:Contains: coat protein gp20; coat protein gp70

C:Species: simian AIDS retrovirus SRV-1

A:Note: host Macaca mulatta (rhesus macaque)

C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 16-Jul-1999

C:Accession: A04000

R:Power, M.D.; Marx, P.A.; Bryant, M.L.; Gardner, M.B.; Barr, P.J.; Luciw, P.A.

Science 231, 1567-1572, 1986

A:Title: Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome

A:Reference number: A94711; MUID:86151668; PMID:3006247

A:Accession: A04000

A:Molecule type: DNA

A:Residues: 1-587 <POW>

A:Cross-references: GB:M11841; NID:G334746; PIDN:AAA47733.1; PID:G334750

C:Comment: The env polyprotein contains coat protein gp70 and coat protein gp20; however

C:Genetics:

A:Gene: env

C:Superfamily: type C retrovirus env polyprotein

C:Keywords: capsid protein; coat protein; glycoprotein; polyprotein

F:120,237,266,271,277,280,295,308,322,328,340,358,488,585/Binding site: carbohydrate (As

Query Match 20.9%; Score 599; DB 1; Length 587;

Best Local Similarity 30.2%; Pred. No. 7.5e-37;

Matches 164; Conservative 86; Mismatches 191; Indels 102; Gaps 20;

QY 24 PCRC---WTSSSPYQELFMQPGNIDAPSYSLKSGTPTFAHTH--- 67

Db 41 PDCAGGVSPSPNSUT-----TVSCSTYTAYSVNSLKWOCVSTPTTASPTHGSC 93

QY 68 -----MPRNCYHSATLCMHANTHWYTGMI---NPS----- 95

Db 94 PSQCNSQSYDSVHATCYNHYQOCTIGNKTYLTATMIRDKSPSSGDGNNVTILGNQNLLI 153

QY 96 --CP--GGLGVTVCTYQTQGMDSGGGVQQAQAEKHVKVISQLTRVHGTSSPYKGLDLS 152

Db 154 AGCPENKKGVVCCWNSQPSVHMSDGGGPGQDKVREIIVNKKFEEHLKSLFELSHPALP 213

QY 153 KLH--ETLRTHRLVSLFN--TTLTGLHEVSAQN--PTNCWICLPLNFRPVYIPVPEQW 206

Db 214 EARGKEKIDAHT-----FDLLATVHSLNVSSORQLAEDCWCLC---RSGDPVPLALPY 264

QY 207 NNFSTEINT-----TSVLVGP--LYSNLEITHSNLTGCV--KFSNTYIT---T 247

Db 265 DNTSCSNSTFFNCSNCSCLITPPFLVQPFNFTHS-----VCLYADYQNSDIDVGLAGFT 321

QY 248 NSQCIRWTPPTQIVCLPSGIFVCGTS--AYRCLNGSSEMCFLSFLVPPMTIYEQD-- 304

Db 322 NCSSVINISKPSPLCAPNSVFCGNNKAYTLPNTWTCVLAATLLPDIDIPGSEPV 381

QY 305 ---LYSYVISKPRNKRVPILPFVIGAGVLGALGTGIGITTSQFYKLSQELNGDMERV 361

Db 382 PIPADHFLGRPK-RAIQFIPLVIGLITTAVSTGTAGLVSLTQYTKLSHQLISDVQAI 440

QY 362 ADSLVTLQDQLNSLAAYVLQNRRLDLLTAERGTCFLFLGEECCYVYVQSGIVTEKVEI 421

Db 441 SSTQDQDQVDSLAEVVLQNRGLDILLTAEQGGICLALQEKCCFYANKSGIVRDKIKNL 500

QY 422 RDRIQRAEELRNTGPMGLLSQWMPWILPFLGPLAAIILLLPFCPIFNLLVNFVSSRIE 481

Db 501 QDDLEKRRKQLIDNPFWTGFHGLLPYVWPLIGPLLCILLVLSFGPIIFNKLMTFIKHQIE 560

QY 482 AVK 484

Db 561 SIQ 563

RESULT 3

VCFVAS

env polyprotein - avian spleen necrosis virus

N:Alternate names: coat polyprotein

N:Contains: coat protein gp22; coat protein gp73

C:Species: avian spleen necrosis virus

C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 01-Mar-1996

C/Accession: A38212
 R/Kewalramani, V.N.; Panganiiban, A.T.; Emerman, M.
 J. Virol. 66, 3026-3031, 1992
 A/Title: Spleen necrosis virus, an avian immunosuppressive retrovirus, shares a receptor
 A/Reference number: A38212; MUID:92219390; PMID:1313915
 A/Accession: A38212
 A/Molecule type: DNA
 A/Residues: 1-567 <KEW>
 A/Cross-references: GB:M87666
 C/Genetics:
 A/Gene: env
 C/Superfamily: type C retrovirus env polyprotein
 C/Keywords: coat protein; glycoprotein; polyprotein; transmembrane protein
 F:1-401/Domain: extracellular #status predicted <EXT>
 F:1-397/Product: coat protein gp73 #status predicted <CPI>
 F:363-379/Region: hydrophobic #status predicted
 F:394-397/Region: cleavage processing #status predicted
 F:398-567/Product: coat protein gp22 #status predicted <CP2>
 F:402-418/Domain: transmembrane #status predicted <TM1>
 F:419-567/Domain: intracellular #status predicted <INT>
 F:245,274,306,328,335/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 20.7%; Score 593.5; DB 1; Length 567;
 Best Local Similarity 33.5%; Pred. No. 1.9e-36;
 Matches 154; Conservative 75; Mismatches 176; Indels 55; Gaps 13;

QY 66 TEMPRNCHSATLCHMANTHWY-----GWINPSCPGGLGVTVCTYFT 110
 DB 121 TOMSTCYETKQETLLGKYFTTAILQTKLGSYEDGPNKLIQASCTGVGKVCWDPA 180
 QY 111 QTGMSDGGVQDQAREKHREVISQLTRVHTGTSYKGLDLSKLHETLRTHLVSLFNT 170
 DB 181 PVYVSDGGPTDMIRESVRELEEIRHSYVPSVQVHPLALP-----SRGVLDLPQ 232
 QY 171 T---LTGLHEV-SQNP---TNCWICLPNFRPVYVSIPEQWNNPSTEINTSVLGL 223
 DB 233 TSDILEATHQVFNATNPKLAENCLWCLTGTPIPAALP-----TNGNVLDTGNCLSLFF 287
 QY 224 VSN-----LEIT---HTSNITCVKFSNTVYTNQICRWTPPTQ-----IVCLPSGIF 269
 DB 288 GCNPPSGIDVSCVAGEADNTGTPGVGVHP-TNCTSIQEVNTSOMGNLTLCPPPGHV 346
 QY 270 FVCGTS-AYCLNGSSESMCFSLFPPMTIYEQD---LYSVISKPNKR-VPLPFPV 324
 DB 347 FVCGNNMAYTALPNKWTGILASIVDIIISGEEPIPLFSTIETARRHKRAVQPIPL 406
 QY 325 IGAGVLGALGTGGITTSQFYKLSQELNGDMERVADSLVTLQDQLNSLAAVLQNR 384
 DB 407 VLGISATLAGTGLGVSVHTKLSNQLIEDVQALSGTINDLQDQDLSAEVVLQNR 466
 QY 385 ALDLTLAERGTCFLGEECCYVNSGIVTERKVEIRDRIQRAEELRNTGFWGLLSQ 444
 DB 467 GLDLLTAEOGGICLALQEKCCFYANKSGIVRDKIRKLQEDLIERKALYDNPWLSGLNGF 526
 QY 445 MPWILPLGLPLAAILLLLPFCIFNLLVNFVSSREAVK 484
 DB 527 LPVLLPLGLPLGLIPLTLGCPIMKTLTRIHDKIQAVK 566

RESULT 4
 VCLJMP
 env polyprotein - Mason-Pfizer monkey virus
 N/Alternate names: coat polyprotein
 N/Contains: coat protein gp20; coat protein gp70
 C/Species: Mason-Pfizer monkey virus
 C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
 C/Accession: D25839
 R/Sonigo, P.; Barker, C.; Hunter, E.; Wain-Hobson, S.
 Cell 45, 375-385, 1986
 A/Title: Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type 2
 A/Reference number: A90878; MUID:86189951; PMID:2421920
 A/Accession: D25839
 A/Molecule type: DNA

A/Residues: 1-586 <SON>
 A/Cross-references: GB:M12349; NID:g334702; PIDN:AAA47712.1; PID:g334705
 A/Experimental source: clone 6A
 C/Genetics:
 A/Gene: env
 C/Superfamily: type C retrovirus env polyprotein
 C/Keywords: capsid protein; coat protein; glycoprotein; polyprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-394/Product: coat protein gp70 #status predicted <GP7>
 F:395-586/Product: coat protein gp20 #status predicted <GP2>
 F:120;237,264,276,291,304,318,324,339,357,487/Binding site: carbohydrate (Asn) (covalent)

Query Match 20.7%; Score 593.5; DB 1; Length 586;
 Best Local Similarity 28.6%; Pred. No. 1.9e-36;
 Matches 156; Conservative 89; Mismatches 192; Indels 109; Gaps 16;

QY 24 PCRCMTSSSPYQEFLLWRMQRGNIDAPSYRSLK-----GTPFTT 63
 DB 41 PCDC-----AGYVSSPINSITVSCSTHTAYSVTNSLKWQCVSTPTTP 85
 QY 64 AMTH-----MPRNCVHSATLCHMANTHWYG----- 89
 DB 86 SNTHIGSCFECNTISYDSVHASCYNHYQCNIKNKTYLTATITGDRTPAIGDGNVPTVL 145
 QY 90 ----XWINPSCPGG-LGVTVCTYFTQTCMSDGGVQDQAREKHREVISQLTRVHTSS 144
 DB 146 GTSENLITAGCPNGKKGQVVCNPSVHISDGGQDQKARDIIVNKKFEELHRSFLPEL 205
 QY 145 PYKGLDLSKLH-EFLRHT-----RLVSLFNTTLTGLHEVSAQNPNTNCWICLPNFRPY 197
 DB 206 SYHPLALPEARKEKIDARTLDLLATVHSLNASQPSLAE-----DCWLCLQSGDPVP 258
 QY 198 VSIPEQ-WNNFSTEINTSVLQGP-LVSNLEIHTSNLTCVKFSNTTYTN-----S 249
 DB 259 LALPYNDILCNFACLSNHSCLPFPFLVQPFNT-DSNCLYAHYQNNSFIDVGLASFT 317
 QY 250 QC-----IRWTPPQIVCLPSGIFVCGTS-AYCLNGSSESMCFSLFPPMTIYEQ 303
 DB 318 NCSSYVNSTASKPSNSLCAPNSSVFCGNKAVTYLPTNWTGSCVLAATLLPDIDILPGS 377
 QY 304 D-----LYSVISKPNKRVPILPVIGAGVLGALGTGIGTITSTQFYKLSQELNGDM 358
 DB 378 EVPPIPAIDHFLGKAK-RAIQILPFLVGLGITTAVSTGAAGLGVSTIYTKLSHQLISDV 436
 QY 359 ERVADSLVTLQDQLNSLAAVLQNRRLDRLTAERGTCFLGEECCYVNSGIVTERKV 418
 DB 437 QAISSTIQDLQDQDLSAEVVLQNRRLDRLTAERGTCFLGEECCYVNSGIVTERKV 496
 QY 419 KEIRDRIQRAEELRNTGFWGLLSQWMPILPFLPLAAIILLLLPFCIFNLLVNFVSS 478
 DB 497 KNLQDLERRRQLDNPFWTSFHGFLPYVMPILLGLPLCLLLVLSFGPIIFNKMTFIKH 556
 QY 479 RIEAVK 484
 DB 557 QIESIQ 562

RESULT 5
 VCLJHD
 env polyprotein precursor - squirrel monkey retrovirus SMRV-H
 N/Alternate names: coat polyprotein
 N/Contains: coat protein gp20; outer membrane protein
 C/Species: squirrel monkey retrovirus SMRV-H
 C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
 C/Accession: D31827
 R/Oda, T.; Ikeda, S.; Watanabe, S.; Hattushika, M.; Akiyama, K.; Mitsunobu, F.
 Virology 167, 468-476, 1988
 A/Title: Molecular cloning, complete nucleotide sequence, and gene structure of the provi
 A/Reference number: A31827; MUID:89073750; PMID:3201749
 A/Accession: D31827
 A/Molecule type: DNA
 A/Residues: 1-575 <ODA>
 A/Cross-references: GB:M23385; NID:g332626; PIDN:AAA66455.1; PID:g332627

QY 13 LLPSELTAPPPCRMTSSPVQEFWMRQPGNIDAPSYRSLSGTGTFTTAHTMPRNC 72
Db 101 IIPS-VQOPGPCCLTIA-----MHAN-THYTGKMINPSCPGGLGVTV 125
QY 73 YHSATLC-----MHAN-THYTGKMINPSCPGGLGVTV 104
Db 126 YKAECECTLLGKTYFTAILOKTKLGSYEDGPNKLLQASCTGIMWTSMLGRCP-----CV 180
QY 105 CWTYFTQGMDDGGVQDQAREKHKEVISOITRVHGTSSPYKGLDLSKLHETRLTHRL 164
Db 181 C-----LDGGGPTDRF-GRICAEGLERIRHSVPSVQYHPLALPR-----PRG 222
QY 165 VSLFNWT---LTGLHEV-SANP---TNCWICLPLNFRPYVSIIPVQMNPFSTEINTTS 217
Db 223 VDLDPQTSDDILEATHOVNLNATNPOLAENCWLCMTLG-----TQSPQSRMMAMS-----LS 273
QY 218 VLVGLVLENLEITHSN---LTC-----VKFSNTTNTTNSQCIIRWVTPPTOI 261
Db 274 MEIAVLASGATHRVNRCQLCREADNRGTPIGVYVHFTNCTSIQSLTRVYIELRD 333
QY 262 VCLPSGIFVCGTSAYRCLNGSSSMCFSLFLVPMPTIYTEQD---LYSVYISKPRNKR- 317
Db 334 VYLRVWYLCVEQAHYATLPKNWIGLCILASIVPDMISIIPGEEPIPLPSIEYTAGRHKA 393
QY 318 VPIPLFVIGAGVLGALGTGIGITTSOTPYKLSQELNGDMERVADSLVTLQDOLNSLAA 377
Db 394 VOFIPLLVGLGTATLGGTGLGVSVTHYKLSNQLIEDVQALSGTINDLQDQIDSLAE 453
QY 378 VYQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKVEIRDRIRQRAEELRNTGP 437
Db 454 VYQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKVEIRDRIRQRAEELRNTGP 513
QY 438 WCLLSQWMPWLPFLGPIPLAAILLLFGPCIFNLLNFRVSVRIEAVK-LQNEPKQSKTK 496
Db 514 WNLNGFLPYLPLSLGFLGLFILTLPGLPCIRKTLRIHDKIQGKNRPRISPAVQATP- 572
QY 497 IYRRPLDRPAPRSQV 512
Db 573 -----NRDGPVRSNV 582

RESULT 8
env polyprotein - feline endogenous virus RD114 (fragment)
A:Reference number: S12812
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 20-Sep-1999
C:Accession: S12812
R:Moehring, R.
submitted to the EMBL Data Library, February 1990
A:Reference number: S12812
A:Accession: S12812
A:Molecule type: genomic RNA
A:Residues: 1-321 <MOE>
A:Cross-references: EMBL:X51930; NID:g61676; PID:CAA36196.1; PID:g61677
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: polyprotein

Query Match 16.2%; Score 465.5; DB 2; Length 321;
Best Local Similarity 34.4%; Pred No. 3.3e-27;
Matches 105; Conservative 62; Mismatches 105; Indels 33; Gaps 7;

QY 185 NCWICLPLNFRPYVSIIPVQMNPFSTEINTTSVLVGLVSNLITHTSNLTCTVKFNTT 244
Db 24 SCQIIPLLVQP-----MDFNSNSSLSPFIND---TEQIDLGAVTFNCT 66
QY 245 YTNNSQCIRWVTPPTQIVCLPSGIFVCGTS-AYRCLNGSSSMCFSLFLVPMPTIYTEQ 303
Db 67 SVAN-----VSSP-----LCALNGSVLCNMAVYLPQNTWRCVQSLPDDINDPGD 117
QY 304 D-----LYSVYISKPRNKRVPILPFGVIGALGTGIGITTSOTPYKLSQELNGDM 358
Db 118 EBPVPIADHYTHREK-RAVQPIPLAGLIGITAAFTTCATGLGVSVTQYTKLSHQLISDV 176

QY 359 ERVADSLVTLQQLNSLAAVVLQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKV 418
Db 177 QVLSGTIQDQLQDQVDSLAEVVLQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKV 236
QY 419 KEIRDRIQRAEELRNTGPWGLLSQWMPWLPFLGPIPLAAILLLFGPCIFNLLNFRVSV 478
Db 237 RILQELQKRRSLATNPWTGLQGLPFLYLLPLUGLPLLLILLFPGPCVFRSLMAFIND 296
QY 479 RIEAV 483
Db 297 RLNVV 301

RESULT 9
env polyprotein precursor - Rauscher mink cell focus-forming virus
N:Contains: coat protein p12E; coat protein p15E; knob protein gp70
C:Species: Rauscher mink cell focus-forming virus
C:Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03990
R:Vogt, M.; Haggblom, C.; Swift, S.; Haas, M.
J. Virol. 55, 184-192, 1985
A:Title: Envelope gene and long terminal repeat determine the different biological properties of Rauscher mink cell focus-forming virus
A:Reference number: A93011; MUID:85237696; PMID:4009793
A:Accession: A03990
A:Molecule type: DNA
A:Residues: 1-640 <VOG>
A:Cross-references: GB:M10100; NID:g332068; PIDN:AAA46528.1; PID:g332070
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-443/Product: knob protein gp70 #status predicted <XP>
F:444-640/Product: coat protein p15E #status predicted <PFE>
F:444-623/Product: coat protein p12E #status predicted <PTE>
F:43,58,300,332,339,372,408,576/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.8%; Score 424; DB 1; Length 640;
Best Local Similarity 30.0%; Pred No. 1e-23;
Matches 111; Conservative 63; Mismatches 144; Indels 52; Gaps 10;

QY 160 THTRVLSFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSIIPVQMNPFSTEINTSVL 219
Db 284 TGRLLNLVDGAYQALNLTSPDKTQECWLCVAEPYVYEGVAVLGTYSNHTSAPTNCVA 343
QY 220 VGLVSNLEIT-----HTSNLTCTVKFSNTTNTTNSQCIIRWVTPPTQIVCLPSGIF 269
Db 344 SQHKLTSEVTGGGLCIGTVPKTHQALC-----NTLTKNK-----GSYLVAAGYT 391
QY 270 FVCGTSAYRCLNGS-----SESMCFSLFLVPMPTIYTEQDLYSVYISKPRNKRVP 322
Db 392 WACNTGLTCLSATVNLRTDYCVLWELWPRVTHPPSVYVYVYVYVYVYVYVYVYVYVYV 451
QY 323 FVIGAGVLGALGTGIGIT---STQPYKLSQELNGDMERVADSLVTLQDOLNSLAAY 379
Db 452 LLLGLTMMGIAAGVGTGTALVATQOOLAAVQOOLAAVQOOLAAVQOOLAAVQOOLAAV 511
QY 380 LQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKVEIRDRIRQRAEELRNTGP 438
Db 512 LQNRRLDILLTARGGTCLFLGEECCYVYVNSQSGIVTEKVEIRDRIRQRAEELRNTGP 571
QY 439 -GLLSQWMP-----LPLFGLPAAIILLFGPCIFNLLNFRVSVRIEAVK----- 484
Db 572 EGLFNR-SPWFTTLISTINGPLIILLFPGPCVFRSLMAFIND 630
QY 485 ----LQWEPK 490
Db 631 QLRPLEYEPQ 640

RESULT 10
T01381

env protein - murine leukemia virus
 C:Species: murine leukemia virus
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Jun-2000
 C:Accession: T01381
 R:Tumab, K.M.; Poszay, J.M.; Avidan, N.; Ksiazek, S.J.; Overmeyer, B.; Blank, K.J.; Fry
 Virology 192, 587-595, 1993
 A:Title: Loss of antigenic epitopes as the result of env gene recombination in retrovirus
 A:Reference number: Z14313; MUID:93134803; PMID:7678475
 A:Accession: T01381
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-676 <TUM>
 A:Cross-references: EMBL:SS3043; NID:G263651
 C:Genetics: env
 C:Gene: env
 C:Superfamily: type C retrovirus env polyprotein

Query Match 14.6%; Score 419; DB 2; Length 676;
 Best Local Similarity 25.0%; Pred. No. 2.5e-23;
 Matches 153; Conservative 70; Mismatches 214; Indels 174; Gaps 25;

QY 20 TAP-PPCRMTSS-----SPQEFLEWQRPNGN---IDAPSYSLSGKTPTFTA 64
 DB 100 SAGPPC-CSGSSGNIAGCARDNEPLTSLTPRCNTAGNRLKLDQVTHKS-SEGFFVCPG 157
 QY 65 HTMPR---NC-----YHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTYFT-QTGM 114
 DB 158 -SHRPREAKSCGPDSPFCASWCETTRGVW-----KPSSP-----WDYITVDNNL 203
 QY 115 SDGGGVQDQAREKHVKVEIQLTR-----VHGTSPPYKGLDL----- 151
 DB 204 TSNQAVQCKDNKWCNPLAQFTNAGQVTSWITGHYWGRLVLSQDPLGFLGRLSYQ 263
 QY 152 -----SKLHETL-----RTHTR 163
 DB 264 NLGPRIPIGNPVLADQLSPFLNPLPKPAKSPASSPTLISPSPTPTPPVGTGDR 323
 QY 164 LVSLPNTTLGLHEVSAQNFTNCWICLPLNFRPYVSIPIVPEQWNNFSTEINTTSVLVGLP 223
 DB 324 LNLVQAGYQALNLTNPKTQECWLCVLSGPPYEGVAV-----LGTSNHTSA----- 372
 QY 224 VSNLEITHNLTCVKSNTTYTNSQCIRWVTPPTQIVC-----LPSGIF 269
 DB 373 PANCVASOHLTL-----SEVTGRLCIGTVPKTHQALCNTTLKAGKSYVLVAFTGM 427
 QY 270 FVCGTSAYRCLNGS-----SESMCFSLPLVPMIYIEQDLYSVISKPRNKRVP1---LP 322
 DB 428 WACNTGLTPCLSATVLRNTDYCVLVELWPRVTHYHPPSYVYQFEXSHRHKREPVSULTA 487
 QY 323 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQLNSLAUV 379
 DB 488 LLLGLTMMGGIAAGVGTGTTALVATQFQQLHAAVQDDLKEVEKSIITNLEKSLTSLSEV 547
 QY 380 LQNRALDLLTAERGGTCLFLGEECCYVYVNSQGIIVTEKVKIIRDIORRAEELRNTGPG 439
 DB 548 LQNRGLDLLFLKGGGLCAALKKECCFYADHTGLVRDSMAKRLRSOKLFESSQGW- 506
 QY 440 LLSQW---PW-----ILPFLGPLAAIILLFPGPCIFNLLVNFVSSRIEAVK----- 484
 DB 607 -FEGWFRSPWFTTLISTINGPLIILLLLFPGCILNRLVQVQKDRISVQALVTOY 665
 QY 485 -----LQWEPK 490
 DB 666 HQLKPLEYEPQ 676

RESULT 11
 VCVWEM

env polyprotein - Moloney murine leukemia virus
 N:Contains: knob protein gp70; R protein; spike protein p15E
 C:Species: Moloney murine leukemia virus
 C:Date: 01-Sep-1981 #sequence_revision 24-Sep-1981 #text_change 16-Jul-1999
 C:Accession: A93265; A93235; A93848; A03983

R:Shinnick, T.M.; Lerner, R.A.; Sutcliffe, J.G.
 Nature 293, 543-548, 1981
 A:Title: Nucleotide sequence of Moloney murine leukaemia virus.
 A:Reference number: A93265; MUID:82035843; PMID:6169994
 A:Accession: A93265
 A:Molecule type: genomic RNA
 A:Residues: 1-665 <SH>
 A:Cross-references: GB:J02255; GB:J02256; GB:J02257; GB:M76668; NID:G331934; PIDN:AAB599
 A:Experimental source: clone pMLV-1
 R:Sutcliffe, J.G.; Shinnick, T.M.; Verma, I.M.; Lerner, R.A.
 Proc. Natl. Acad. Sci. U.S.A. 77, 3302-3306, 1980
 A:Title: Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replic
 A:Reference number: A93848; MUID:81013872; PMID:6251454
 A:Accession: A93848
 A:Molecule type: DNA
 A:Residues: 484-662, 'CEF' <SUT2>
 A:Experimental source: provirus, clone pMLV-201
 C:Genetics:
 C:Gene: env
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
 F:34-469/Product: knob protein gp70 #status predicted <KN>
 F:470-649/Product: spike protein p15E #status predicted <SPK>
 F:650-665/Product: R protein #status predicted <RPT>
 F:45,199,326,358,398,434/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.6%; Score 418.5; DB 1; Length 665;
 Best Local Similarity 30.2%; Pred. No. 2.7e-23;
 Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THTRLVSLPNTTLGLHEVSAQNFTNCWICLPLNFRPYVSIPIVPEQWNNFSTEINTSVL 219
 DB 310 TENLLNLVDCAYQALNLTSPDKTQECWLCVAGPPYEGVAVLGTYSNHTSAPANCVA 369
 QY 220 VGPLVSNLEIT-----HTSNLTCVKSNTTYTNSQCIRWVTPPTQIVCLPSGIF 269
 DB 370 SQHKLTLSEVGTGQGLCIGAVPKTHQALC-----NTQTSR-----GSYILVAPGTGM 417
 QY 270 FVCGTSAYRCLNGS-----SESMCFSLPLVPMIYIEQDLYSVISKPRNKRVP1---LP 322
 DB 418 WACSTGLTPCLSATVLRNTDYCVLVELWPRVTHYHPPSYVYVGLPERSNRHKREPVSULTA 477
 QY 323 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQLNSLAUV 379
 DB 478 LLLGLTMMGGIAAGVGTGTTALVATQFQQLHAAVQDDLKEVEKSIITNLEKSLTSLSEV 537
 QY 380 LQNRALDLLTAERGGTCLFLGEECCYVYVNSQGIIVTEKVKIIRDIORRAEELRNTGPGW- 438
 DB 538 LQNRGLDLLFLKGGGLCAALKKECCFYADHTGLVRDSMAKRLRSOKLFESSQGW 597
 QY 439 -GL-SQWMPW-----ILPFLGPLAAIILLFPGPCIFNLLVNFVSSRIEAVK 484
 DB 598 EGLFNR-SPWFTTLISTINGPLIILLLLFPGCILNRLVQVQKDRISVQ 647

RESULT 12
 S70395
 env polyprotein - Friend murine leukemia virus (strain FB29)
 N:Contains: knob protein gp76; R protein; spike protein p15E
 C:Species: Friend murine leukemia virus
 A:Variety: strain FB29
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
 C:Accession: S70395
 R:Perryman, S.; Nishio, J.; Chesebro, B.
 Nucleic Acids Res. 19, 6950, 1991

A:Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
 A:Reference number: S70393; MUID:92107687; PMID:1762923
 A:Accession: S70395
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: genomic RNA
 A:Residues: 1-676 <PER>
 A:Cross-references: EMBL:Z11128; NID:G61547; PID:CAA77479.1; PID:G61549
 A:Experimental source: strain FB29
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
 C:Genetics:
 A:Gene: env
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: coat protein; polyprotein; spike protein; transmembrane protein

Query Match 14.4%; Score 414; DB 2; Length 676;
 Best Local Similarity 29.3%; Pred. No. 6e-23;
 Matches 110; Conservative 56; Mismatches 147; Indels 62; Gaps 11;

QY 160 THRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSL 219
 DB 320 TGRLLNLVQAYQALNLTNPDKTQECWLCVSGPPYEGVAV-----LGTYSNHTSA- 372

QY 220 VPLVSNLEITHSNLTCVKFSNTTYTNSOCIRWVTPPTQIVC-----LP 265
 DB 373 ----PANCVASOQKLT-----SEVTGGLCIGVPTKTHOALCNLTAKGSGYILVAP 423

QY 266 SGIFVCGTSAYRCINGS-----SESMCFSLFLVPPMTIYTEQDLYSYVVISKPRNKRVP- 320
 DB 424 TGTWACNTGLTFCISATVLRNTDTCVVLWPRVTVHPSPVYVVSQPEKSHRKRBPVS 483

QY 321 --LPFVIGAGVLGALTGGIGITT---STQFYKLSQELNGDMERVADSLVTLODQNSL 375
 DB 484 LFLALLLGLTWGGTAAGVGTGTALVATQFQQLAAVQDQLKEVEKSIINLEKSLTSL 543

QY 376 AAVLQNRALDLLTAERGTCFLGEBCCYVNSQGIIVTEKVEIRDRIOREBELRNT 435
 DB 544 SEVLQNRGLDLLFLKEGGLCAALKKECCFYADHTGLVRDSMAKLERLSQRKLFESS 603

QY 436 GPWGLLSOM-----PW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK----- 484
 DB 604 QGW--PEGWFRSPWFTLTISTINGPLIILLFGPCILNRLVQFKDRISVVQALV 661

QY 485 -----LQMEPK 490
 DB 662 TQOYHQLKPLEYEQ 676

RESULT 13
 A43491
 env polyprotein - T3651/B murine leukemia virus (strain T1223/B)
 C:Species: T3651/B murine leukemia virus
 C:Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 20-Apr-2001
 R:Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand, J
 A:Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-in
 A:Reference number: A43491; MUID:91253260; PMID:2042396
 A:Accession: A43491
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-689 <BAY>
 A:Cross-references: GB:X59002
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 14.4%; Score 413.5; DB 2; Length 689;
 Best Local Similarity 27.8%; Pred. No. 6.7e-23;
 Matches 110; Conservative 67; Mismatches 151; Indels 67; Gaps 12;

QY 163 RLVLNFTLTGLHEVSAQNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGP 222
 DB 332 RLNLVKGAYQALNLTSPDKTQECWLCVSGPPYEGVAV-----LGTYSNHTSA- 381

QY 223 LVSNLEITHSNLTCVKFSNTTYTNSOCIRWVTPPTQIVC-----LPSGI 268
 DB 382 -PANCVASOQKLT-----SEVTGGLCIGAVPKTLQVLCNTTQKTSTGSYLAAPTGT 435

QY 269 FVVGTSAYRCINGS-----SESMCFSLFLVPPMTIYTEQDLYSYVVISKPRNKRVP- 321
 DB 436 IWACSTGLTFCISATVLRNTDTCVVLWPRVTVHPSPVYVVSQPEKSHRKRBPVS 495

QY 322 PFVIGAGVLGALTGGIGITT---STQFYKLSQELNGDMERVADSLVTLODQNSLA 378
 DB 496 ALLGLLTMGGTAAGVGTGTALVATQFQQLAAVQDQLKEVEKSIINLEKSLTSLSEV 555

QY 379 VLQNRALDLLTAERGTCFLGEBCCYVNSQGIIVTEKVEIRDRIOREBELRNTGPW 438
 DB 556 VLQNRGLDLLFLKEGGLCAALKKECCFYADHTGLVRDSMAKLERLSQRKLFESSQGW 615

QY 439 --GLLSQWMPW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVKQMEPKM 492
 DB 616 FEGLFNK-SPWFTLTISTINGPLIILLFGPCILNRLVQFKDRISVVQALV----- 669

QY 493 SKTIYRRPLDRPASPRSDVNDIKGTPPEISAAQ 527

QY 223 LVSNLEITHSNLTCVKFSNTTYTNSOCIRWVTPPTQIVC-----LPSGI 268
 DB 381 -PANCVASOQKLT-----SEVTGGLCIGAVPKTLQVLCNTTQKTSTGSYLAAPTGT 434

QY 269 FVVGTSAYRCINGS-----SESMCFSLFLVPPMTIYTEQDLYSYVVISKPRNKRVP- 321
 DB 435 IWACSTGLTFCISATVLRNTDTCVVLWPRVTVHPSPVYVVSQPEKSHRKRBPVS 494

QY 322 PFVIGAGVLGALTGGIGITT---STQFYKLSQELNGDMERVADSLVTLODQNSLA 378
 DB 495 ALLGLLTMGGTAAGVGTGTALVATQFQQLAAVQDQLKEVEKSIINLEKSLTSLSEV 554

QY 379 VLQNRALDLLTAERGTCFLGEBCCYVNSQGIIVTEKVEIRDRIOREBELRNTGPW 438
 DB 555 VLQNRGLDLLFLKEGGLCAALKKECCFYADHTGLVRDSMAKLERLSQRKLFESSQGW 614

QY 439 --GLLSQWMPW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVKQMEPKM 492
 DB 615 FEGLFNK-SPWFTLTISTINGPLIILLFGPCILNRLVQFKDRISVVQALV----- 668

QY 493 SKTIYRRPLDRPASPRSDVNDIKGTPPEISAAQ 527
 DB 669 -LQOY-----HOLKSIDPEVESRE 688

RESULT 14
 B43491
 env polyprotein - T3651/B murine leukemia virus
 C:Species: T3651/B murine leukemia virus
 C:Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 20-Apr-2001
 R:Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand, J
 A:Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-in
 A:Reference number: A43491; MUID:91253260; PMID:2042396
 A:Accession: B43491
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-689 <BAY>
 A:Cross-references: GB:X59001
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 14.4%; Score 413.5; DB 2; Length 689;
 Best Local Similarity 27.8%; Pred. No. 6.7e-23;
 Matches 110; Conservative 67; Mismatches 151; Indels 67; Gaps 12;

QY 163 RLVLNFTLTGLHEVSAQNPTNCWICLPLNRPVYSIPVPEQWNNFSTEINTTSLVGP 222
 DB 332 RLNLVKGAYQALNLTSPDKTQECWLCVSGPPYEGVAV-----LGTYSNHTSA- 381

QY 223 LVSNLEITHSNLTCVKFSNTTYTNSOCIRWVTPPTQIVC-----LPSGI 268
 DB 382 -PANCVASOQKLT-----SEVTGGLCIGAVPKTLQVLCNTTQKTSTGSYLAAPTGT 435

QY 269 FVVGTSAYRCINGS-----SESMCFSLFLVPPMTIYTEQDLYSYVVISKPRNKRVP- 321
 DB 436 IWACSTGLTFCISATVLRNTDTCVVLWPRVTVHPSPVYVVSQPEKSHRKRBPVS 495

QY 322 PFVIGAGVLGALTGGIGITT---STQFYKLSQELNGDMERVADSLVTLODQNSLA 378
 DB 496 ALLGLLTMGGTAAGVGTGTALVATQFQQLAAVQDQLKEVEKSIINLEKSLTSLSEV 555

QY 379 VLQNRALDLLTAERGTCFLGEBCCYVNSQGIIVTEKVEIRDRIOREBELRNTGPW 438
 DB 556 VLQNRGLDLLFLKEGGLCAALKKECCFYADHTGLVRDSMAKLERLSQRKLFESSQGW 615

QY 439 --GLLSQWMPW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVKQMEPKM 492
 DB 616 FEGLFNK-SPWFTLTISTINGPLIILLFGPCILNRLVQFKDRISVVQALV----- 669

QY 493 SKTIYRRPLDRPASPRSDVNDIKGTPPEISAAQ 527

Db 670 -LTQQY-----HQLKSIDPEVESRE 689

RESULT 15

VCWVGF

env polyprotein - feline leukemia virus (strain Gardner-Arnstein)

N:Contains: knob protein gp70; R protein; spike protein p15E

C:Species: feline leukemia virus

A:Note: host Felis sp. (cat)

C:Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999

C:Accession: A03991

R:Numberg, J.H.; Williams, M.E.; Innis, M.A.

J. Virol. 49, 629-632, 1984

A:Title: Nucleotide sequences of the envelope genes to two isolates of feline leukemia v

A:Reference number: A92996; MUID:84115095; PMID:6319767

A:Accession: A03991

A:Molecule type: DNA

A:Residues: 1-662 <NUN>

A:Cross-references: GB:K01209; NID:G323910; PID:AAA43052.1; PID:G323911

R:Eider, J.H.; McDee, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Grar

J. Virol. 61, 8-15, 1987

A:Title: Localization of neutralizing regions of the envelope gene of feline leukemia vi

A:Reference number: A25982; MUID:87061257; PMID:2431166

A:Contents: annotation; peptide synthesis

A:Note: 27 peptides synthesized

C:Genetics:

A:Gene: env

C:Superfamily: type C retrovirus env polyprotein

C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

F:1-33/Domain: signal sequence #status predicted <SIG>

F:34-465/Product: knob protein gp70 #status predicted <KPG>

F:466-645/Product: spike protein p15E #status predicted <SKP>

F:646-662/Product: R protein #status predicted <RSP>

F:43,58,286,322,327,351,354,394,410,430/Binding site: carbohydrate (Asn) (covalent) #sta

Query Match 14.4%; Score 412.5; DB 1; Length 662;

Best Local Similarity 23.2%; Pred. No. 7.6e-23;

Matches 138; Conservative 89; Mismatches 208; Indels 161; Gaps 21;

QY 44 PG-NIDAPSVRLSKGTPTTAHTM-----PRNCYHSATLCMHANTHYWTGMINPS 95

Db 94 PGVCDQPMRWOORNTFFVCPGHANRKCQGGPQDGCFAVWCCTTGETY 148

QY 96 -----CPGGLGVTCW---TYFTQTGMSDGG----- 119

Db 149 SSWDYITVKGVTOGIVQCSCGCGWCGPCYDKAVHSITGASEGRCNPLILQFTQKGRAT 208

QY 120 -----VQDQ-----AREKHVKEVI 133

Db 209 SWDGPKSWGLRYSYDPTALFSVRQVMTITPPQAMGNLVPDQKPPSRQSIERSV 268

QY 134 SQ-----LTVRHGTSSPYKGLDLSKHLHETLRLVLSLFTTLTGLHEVSAQN 182

Db 269 TPHSQNGGTPGTLVNASIAPLSTPVPASPKEIGTGRDLNLVQGYLALNATDPNR 328

QY 183 PTNCWICLPLNFRPYVPIVPEQWNNFSTEINTSVLVGLVSNLEITHSNLTVCVFSN 242

Db 329 TKDCWLCVSRPPPYEGIAI---LGNYSNQTN-----PPPSCLSIPOH-----KLTI 372

QY 243 TTYTNSQCIRWVTPPTQIVC-----LPSGIFFCVCGTSAYRC-----LNGS 283

Db 373 SEVSGQGLCIGTVPKTHQALCNETOQGHGAHYLAAPNGTYWACNTGLTFCISMAVLNWT 432

QY 284 SESMCFSLFVPPMTIYTEQDLSYVISKPRNKVPI---LPFVIG-----AGVLGALGTG 336

Db 433 SD-FCVLIELWPRVTVHQPEYVYTHFAKARFRREPISLTVALMLGGLTVGGIAAGVGTG 491

QY 337 IGGITTSQFYKLSQELNGDMERVADSLVTLQDLNSLAAVVLQNRALDILLTAERGRT 396

Db 492 TKALIETAQF-RQLQMAHMTDIOALESISALEKSLTSLSEVVLQNRGDLILFLOEGGL 550

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 551 CAALKKECCFYADHTGLVRDNMAKLRLKORQQLFDSQQGW--FEGWFNKSPWFTLLIS 608

QY 450 PFLGLPAAIILLILLLFGPCIFNLNLFVSSRIEAVKLOMEPKWQSKTKIYRPLDRP 505

Db 609 SIMGFLILLILLILLLFGPCILNRLVQFVKDRISVVQALILTQQYQOIQOY--DPRP 662

RESULT 16

A25982

env polyprotein - feline leukemia virus (strain Gardner-Arnstein)

C:Species: feline leukemia virus

C:Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 24-Feb-1995

C:Accession: A25982

R:Eider, J.H.; McDee, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Grar

J. Virol. 61, 8-15, 1987

A:Title: Localization of neutralizing regions of the envelope gene of feline leukemia vi

A:Reference number: A25982; MUID:87061257; PMID:2431166

A:Accession: A25982

A:Molecule type: protein

A:Residues: 1-662 <ELD>

C:Superfamily: type C retrovirus env polyprotein

C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

Query Match 14.4%; Score 412.5; DB 2; Length 662;

Best Local Similarity 23.2%; Pred. No. 7.6e-23;

Matches 138; Conservative 89; Mismatches 208; Indels 161; Gaps 21;

QY 44 PG-NIDAPSVRLSKGTPTTAHTM-----PRNCYHSATLCMHANTHYWTGMINPS 95

Db 94 PGVCDQPMRWOORNTFFVCPGHANRKCQGGPQDGCFAVWCCTTGETY 148

QY 96 -----CPGGLGVTCW---TYFTQTGMSDGG----- 119

Db 149 SSWDYITVKGVTOGIVQCSCGCGWCGPCYDKAVHSITGASEGRCNPLILQFTQKGRAT 208

QY 120 -----VQDQ-----AREKHVKEVI 133

Db 209 SWDGPKSWGLRYSYDPTALFSVRQVMTITPPQAMGNLVPDQKPPSRQSIERSV 268

QY 134 SQ-----LTVRHGTSSPYKGLDLSKHLHETLRLVLSLFTTLTGLHEVSAQN 182

Db 269 TPHSQNGGTPGTLVNASIAPLSTPVPASPKEIGTGRDLNLVQGYLALNATDPNR 328

QY 183 PTNCWICLPLNFRPYVPIVPEQWNNFSTEINTSVLVGLVSNLEITHSNLTVCVFSN 242

Db 329 TKDCWLCVSRPPPYEGIAI---LGNYSNQTN-----PPPSCLSIPOH-----KLTI 372

QY 243 TTYTNSQCIRWVTPPTQIVC-----LPSGIFFCVCGTSAYRC-----LNGS 283

Db 373 SEVSGQGLCIGTVPKTHQALCNETOQGHGAHYLAAPNGTYWACNTGLTFCISMAVLNWT 432

QY 284 SESMCFSLFVPPMTIYTEQDLSYVISKPRNKVPI---LPFVIG-----AGVLGALGTG 336

Db 433 SD-FCVLIELWPRVTVHQPEYVYTHFAKARFRREPISLTVALMLGGLTVGGIAAGVGTG 491

QY 337 IGGITTSQFYKLSQELNGDMERVADSLVTLQDLNSLAAVVLQNRALDILLTAERGRT 396

Db 492 TKALIETAQF-RQLQMAHMTDIOALESISALEKSLTSLSEVVLQNRGDLILFLOEGGL 550

QY 397 CLFLGEECCYVNSGIVTEKVKIIRDIORRAEELRNTGPGWLLSQWM---PW-----IL 449

Db 551 CAALKKECCFYADHTGLVRDNMAKLRLKORQQLFDSQQGW--FEGWFNKSPWFTLLIS 608

QY 450 PFLGLPAAIILLILLLFGPCIFNLNLFVSSRIEAVKLOMEPKWQSKTKIYRPLDRP 505

Db 609 SIMGFLILLILLILLLFGPCILNRLVQFVKDRISVVQALILTQQYQOIQOY--DPRP 662

RESULT 17

VCWVLB

env polyprotein - feline leukemia virus (strain lambda-B1) (fragment)

N:Contains: coat protein gp70; coat protein p15E

C:Species: feline leukemia virus

C>Date: 31-Mar-1999 #sequence_revision 31-Mar-1999 #text_change 16-Jul-1999
C/Accession: A27172
R/Nicolaissen-Strouss, K.; Kumar, H.P.M.; Fitting, T.; Grant, C.K.; Elder, J.H.
J. Virol. 61, 3410-3415, 1987
A/Title: Natural feline leukemia virus variant escapes neutralization by a monoclonal antibody
A/Reference number: A27172; MUID:88036192; PMID:2444714
A/Accession: A27172
A/Molecule type: DNA
A/Residues: 1-662 <NIC>
A/Cross-references: GB:J03448; NID:9323896; PIDN:AAA43048.1; PID:9323897
C/Genetics:
A/Gene: env
C/Superfamily: type C retrovirus env polyprotein
C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-465/Product: coat protein gp70 #status predicted <GPS>
F:466-662/Product: coat protein p15E #status predicted <GPR>
F:43.58,286,322,327,351,354,430/Binding site: carbohydrate (Asn) #status predicted
Query Match 14.3%; Score 411.5; DB 1; Length 662;
Best Local Similarity 27.8%; Pred. No. 9e-23;
Matches 110; Conservative 69; Mismatches 164; Indels 53; Gaps 11;
QY 136 LFRVHGTSPPYKGLDLSKHLHETRLHSLVSLFNTLTGLHEVSAQNPTNCWICLPLNFR 195
Db 282 IFLVNASIAPLSTPVPASPKEIGTGNRLINLVQGYLALNVTNPNTKDCWCLVSRPP 341
QY 196 PVVSIPTVPEQWNNFSTENTTSVLGVLVSLNLEITHTSNLTCKVPSNTTYTNSQIRWV 255
Db 342 YVEGIAV---LQNYNQNTNPPSCISDPOHKLITSEVSG-----QSGCIGTV 385
QY 256 TPPTQIVC-----LPSGIFPVCGTSAYRC-----LNGSSESCFSLFVPP 296
Db 386 PKTHQALCKTKQGHKTHYLAAPSGTYWACNTGLTPCISMAVLNWTSD-FCVLIELWFR 444
QY 297 MLYTEQDLYSVISKPNKRPVI---LPPVIG---AGVLGALTGGIGGTTSTQFYVK 349
Db 445 VTHQPEYVYTHFDKTVLRREPISLTVALMUGGLTVGGIAGVGTYKALLETAPQ-GQ 503
QY 350 LSQELNGDMERVADSLVTLDQNLAAVQLNRRALDLLTAERGTCFLGEECCYVYN 409
Db 504 LQWAMHTDQALEESIISALEKSLTSLSEVLQNRRLDILFLQEGGLCAALKECCFYAD 563
QY 410 QSGIVTEKVEIRDRIQRAELRNTGFWLLSQM---PW-----ILFPLGLAAIILL 462
Db 564 HTGLVRDNNAKLRERLKRQQLFDSQQGW--FECHFNKSPMFTTLTSSINGPLILLIL 621
QY 463 LFGPCIFNLVNFVSSRIEAVKLQWPKQSKTKY 498
Db 622 LFGPCILNRLVQFVDRISVQALILTQYQOIKQY 657
RESULT 18
VCMVPS
env polyprotein precursor - Friend murine leukemia virus (strain PVC-211)
N/Alternate names: coat polyprotein
N/Contains: knob protein gp76; R protein; spike protein p15E
C/Species: Friend murine leukemia virus
C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999
C/Accession: A38210; S35476
R/Masuda, M.; Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.
J. Virol. 66, 2798-2806, 1992
A/Title: Molecular characterization of a neuropathogenic and nonerythrocytotoxic variant
A/Reference number: A38210; MUID:92219364; PMID:1560524
A/Accession: A38210
A/Molecule type: DNA
A/Residues: 1-676 <SPE>
A/Cross-references: GB:M93134; NID:9331898; PIDN:AAA46478.1; PID:9331901
R/Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.; Masuda, M.
Nucleic Acids Res. 20, 3249, 1992
A/Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leukemia virus
A/Reference number: S35474; MUID:92319660; PMID:1620621
A/Accession: S35476

A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-676 <REM>
A/Cross-references: EMBL:M93134; NID:9331898; PIDN:AAA46478.1; PID:9331901
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
C/Genetics:
A/Gene: env
C/Superfamily: type C retrovirus env polyprotein
C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-676/Product: env polyprotein #status predicted <ENV>
F:36-619/Domain: extracellular #status predicted <EXT>
F:36-479/Product: knob protein gp76 #status predicted <KGP>
F:476-479/Region: cleavage processing #status predicted
F:480-659/Product: spike protein p15E #status predicted <SPP>
F:486-502/Region: hydrophobic #status predicted
F:620-636/Domain: transmembrane #status predicted <TM1>
F:637-676/Domain: intracellular #status predicted <INT>
F:660-676/Product: R protein #status predicted <RPT>
F:46,202,336,368,408,444/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 14.3%; Score 410; DB 1; Length 676;
Best Local Similarity 29.2%; Pred. No. 1.2e-22;
Matches 111; Conservative 58; Mismatches 139; Indels 72; Gaps 12;
QY 160 THRLVSLFNTLTGLHEVSAQNPTNCWICLPLNFRPYVISIPVPEQWNNFSTENTSVL 219
Db 320 TGDRLNLVQAGYQALNLTNPDKTCWCLCL-VSAPPYYE----- 358
QY 220 VGPLVSNLEITHTS-NLTCV-----KFSNTTYTNSQCIRWVTPPTQIVC- 263
Db 359 -GVAVLTGYSNHTSAPANCSSQSHKLTLSVETGGLCIGTVKTHQALCNTLTKTGKS 417
QY 264 ----LPSGIFPVCGTSAYRCNLS-----SESMCFSLFVPPMTIYTEODLYSVISKPN 315
Db 418 YVLVAPAGTVMWACNTGLTPTCLSATVLNRTDYCVLWELWPRVTPYHPSPYVISOFEKSYRH 477
QY 316 KRYPI---LPPFVIGAGVLGALTGGIGIT---STQFYKLSQELNGDMERVADSLVTQ 369
Db 478 KRFPVSLTALLUGGLTGMGIAAGVTGTALTAVTQFQQLHAAVQDDLKEVEKSINTLE 537
QY 370 DQNLAAVQLNRRALDLLTAERGTCFLGEECCYVYNQSGIVTEKVEIRDRIQRA 429
Db 538 KSLTSLSEVLQNRRLDILLFLKEGGLCAALKECCFYADHTGLVDRDSMAKLERLTQRQ 597
QY 430 EELRNTGFW---GLLSQWMPW-----ILFPLGLAAIILLILFGPCIFNLVNFVSSRIEAV 483
Db 598 KLFESSQGWFEGLFNR-SPMFTTLTSTMGPLILLILLFGPCILNRLVQFVDRISVV 656
QY 484 K-----LQWEPK 490
Db 657 QALVLTQYHQLKPLEVEPQ 676
RESULT 19
VCMVSS
env polyprotein precursor - feline sarcoma virus (strain SM)
N/Contains: coat protein gp70; coat protein p15E
C/Species: feline sarcoma virus
A/Note: host feline silvestris catus (domestic cat)
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
C/Accession: A33741
R/Guilhot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
Virology 161, 252-258, 1987
A/Title: Nucleotide sequence analysis of the LTRs and env genes of SM-FesV and GA-FesV.
A/Reference number: A33741; MUID:88044502; PMID:2823466
A/Accession: A33741
A/Molecule type: DNA
A/Residues: 1-645 <GUI>
C/Genetics:
A/Gene: env
C/Superfamily: type C retrovirus env polyprotein
C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein;

Qy

Qy
434 NTGPWGLLSQWM---PW---ILPFLGPLAAIILLLLFGPCIFNLL

Qy 434 NTGPWGLLSQWM---PW---ILPFLGPLAAIILLFLFGPCIFNLLVNFVSSRIEAVKLQ

Db 568 SQQGW--PEGWFKSPWFTTLLISSINGPLILLILLFGPCILNRLVQVKDRISVQAL 625

QY 487 MEPRKQSKTKIYRRPLDRP 505

Db 626 ILTQYQOIQY--DPDRP 642

RESULT 22

VCWVFS

env polyprotein - mink cell focus-forming virus

N/Alternate names: coat polyprotein

N/Contains: knob protein gp70; R protein; spike protein p15E

C/Species: mink cell focus-forming virus

C/Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 23-Aug-1997

C/Accession: A31668

R/Chattopadhyay, S. K.; Baroudy, B. M.; Holmes, K. L.; Fredrickson, T. N.; Lander, M. R.; Mon

Viology 168, 90-100, 1989

A/Title: Biological and molecular genetic characteristics of a unique MCF virus that is hi

A/Reference number: A31668; MUID:89085614; PMID:2535909

A/Accession: A31668

A/Molecule type: DNA

A/Residues: 1-636 <CHA>

C/Genetics:

A/Gene: env

C/Superfamily: type C retrovirus env polyprotein

C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

F/1-30/Domain: signal sequence #status predicted <SIG>

F/31-440/Product: knob protein gp70 #status predicted <GP7>

F/441-620/Product: spike protein p15E #status predicted <P1E>

F/621-636/Product: R protein #status predicted <RPT>

F/43,58,297,329,369,405/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.2%; Score 406.5; DB 1; Length 636;

Best Local Similarity 29.6%; Pred. No. 28-22;

Matches 104; Conservative 64; Mismatches 144; Indels 39; Gaps 9;

QY 160 THRLVSLFNTLTGLHEVSAQNPNCWICLPNFRPVYSIPVPSQWNNFSEINTTSL 219

Db 281 TGRLLNLVDGAYQALNLSPTKQECWCLVAGPYEGVALGTYSNHSAPANCVA 340

QY 220 VGPLVSNLEIT--HTSNLTVCVKNFTNTYTNSQCIKWTPPTQIVC 269

Db 341 SQKHLTSEVTGGCLCVGAVPKTHQALC---NTTKTSD-----GSIYLAAPAGTI 388

QY 270 FVCGTSAYCLNGS-----SESFCFLSFLVPPMTIYEQDLYSYVISKRNKRVPI--LP 322

Db 389 WACNTGLTCLSTVLNLTDDYCVLVELWPKVTHYSPDYVYTFQFPGARFRPEPVSULTA 448

QY 323 FVIGAGVLGALGTGIGGITT---STOFYKLSQELNGMDMERSVADSLVTIQDQLNSLA 379

Db 449 LLLGLTGGIAGVGTGTTALVATQFQQLQAAVHNDLKEVKSITNLEKSLTSLSEVA 508

QY 380 LQNRALDILLTARGTCLFLGECYVYVNSQGIIVTEKVEIRDRIOARRAEELRTGWP- 438

Db 509 LQNRGLDILLFLKEGGLCAALKECCFYADHTGLVRDSMAKLRLNQRQLFESGQGW 568

QY 439 -GLSLQWMPW----ILPFLGLPLAAIILLFGPCIFNLVFNVSRIEAVK 484

Db 569 EGLFNR-SPWFTTLLISTINGPLILLILLFGPCILNRLVQVKDRISVQ 618

RESULT 23

T10533

env polyprotein precursor - feline leukemia virus (strain FeLV-FAIDS)

N/Contains: env protein gp70; env protein p15E

C/Species: feline leukemia virus

C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C/Accession: T10533

R/Donahue, P. R.; Hoover, E. A.; Beltz, G. A.; Riedel, N.; Hirsch, V. M.; Overbaugh, J.; Mul

J. Virol. 62, 722-731, 1988

A/Title: Strong sequence conservation among horizontally transmissible, minimally pathog

A/Reference number: Z17078; MUID:88119207; PMID:2828667

A/Accession: T10533

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-642 <DON>

A/Cross-references: EMBL:ML8247; NID:G323904; PIDN:AAA93093.1; PID:G323906

C/Superfamily: type C retrovirus env polyprotein

Query Match 14.2%; Score 406.5; DB 2; Length 642;

Best Local Similarity 29.0%; Pred. No. 2e-22;

Matches 110; Conservative 65; Mismatches 149; Indels 55; Gaps 13;

QY 160 THRLVSLFNTLTGLHEVSAQNPNCWICLPNFRPVYSIPVPSQWNNFSEINTTSL 219

Db 286 TGRLLNLVQGYTALNATDNPKTKDCWCLVSRPPYEGIAI---LGNYSQNTN----- 337

QY 220 VGPLVSNLEITHTSNLTVCVKNFTNTYTNSQCIKWTPPTQIVC-----LP 265

Db 338 --PPSCLSIPQH-----KLITSEVSGQGLCIGTVPKTHQALCNKTKQOCHTGAHYLAAP 389

QY 266 SGFFVCGTSAYRC-----LNGSESMCFSLFVPPMTIYEQDLYSYVISKRNKRVPI 320

Db 390 NGTYWACNTGLTFCISMAVLNWTSD-FCVLIELWPRVYTHOPEYVYTHFAKAVFRREPI 448

QY 321 --LPFVIG---AGVLGALGTGIGGITTSTQFYKLSQELNGMDMERSVADSLVTIQDQLN 373

Db 449 SLTVALMGLTGVGGIAGVGTGKTALLETAP-RQLQAMHTDIOALEESISALEKSLT 507

QY 374 SLAAVLQNRALDILLTARGTCLFLGECYVYVNSQGIIVTEKVEIRDRIOARRAEELR 433

Db 508 SLSEVQLNRRGLDILLFLQEGGLCAALKECCFYADHTGLVRDNMAKLRLKORQLFD 567

QY 434 NTGSPWGLLSQW---PW----ILPFLGLPLAAIILLFGPCIFNLVFNVSRIEAVKIQ 486

Db 568 SQQGW--FEGWFKSPWFTTLLISSINGPLILLILLFGPCILNRLVQVKDRISVQAL 625

QY 487 MEPRKQSKTKIYRRPLDRP 505

Db 626 ILTQYQOIQY--DPDRP 642

RESULT 24

VCWVEX

env polyprotein - AKV murine leukemia virus

N/Contains: knob protein gp76; R protein; spike protein p15E

C/Species: AKV murine leukemia virus

C/Date: 05-Apr-1983 #sequence_revision 03-Aug-1984 #text_change 16-Jul-1999

C/Accession: A92995; A93448; A03984

R/Herr, W.

J. Virol. 49, 471-478, 1984

A/Title: Nucleotide sequence of AKV murine leukemia virus.

A/Reference number: A92995; MUID:84115072; PMID:6319746

A/Accession: A92995

A/Molecule type: genomic RNA

A/Residues: 1-669 <HEE>

A/Cross-references: GB:J01998; GB:J01999; GB:K00016; GB:K00017; GB:K00018; GB:K01394; NI

R/Herr, W.; Corbin, V.; Gilbert, W.

Nucleic Acids Res. 10, 6931-6944, 1982

A/Title: Nucleotide sequence of the 3' half of AKV.

A/Reference number: A93448; MUID:83090450; PMID:6294621

A/Accession: A93448

A/Molecule type: DNA

A/Residues: 1-34,'R',36-462,'K',464-591,'K',593-669 <HE2>

C/Genetics:

A/Gene: env

C/Superfamily: type C retrovirus env polyprotein

C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

F/1-31/Domain: signal sequence #status predicted <SIG>

F/32-470/Product: knob protein gp76 #status predicted <KNB>

F/471-650/Product: spike protein p15E #status predicted <SPK>

F/651-669/Product: R protein #status predicted <RPT>

F/43,199,327,359,399/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.2%; Score 406.5; DB 1; Length 669;

Best Local Similarity 29.5%; Pred. No. 2.2e-22;

| | |
|---|--|
| Matches 104; Conservative 60; Mismatches 141; Indels 47; Gaps 10; | |
| Qy 163 RLVSLENTTLTGLHEVAQNPTNCWICLPNFRPYVSIPIVPEQWNNFSTEINTTSVLGP 222 | |
| Db 314 RLNLVKGACALNLTSPDKTQECWCLVSGPPYEGVAV-----LGTYSNHTSA---- 363 | |
| Qy 223 LVSNLEIHTSNLTCVKFSNTYTTNSQCIRWVTPPTQIVC-----LPSGI 268 | |
| Db 364 -PANCVASQHKLT-----SEVTGQGLCTGAVPKTHQVLCNTTOKTSDGSYYLAAPTGT 417 | |
| Qy 269 FVCGTSAYRCLNGS-----SESMLFLSFLVPPMTIYTEQDLYSVVISKPNKRVPI---L 321 | |
| Db 418 TWACSTGLTFCISTIIDLTDCVILWELWFRVTHSPSYVYHQFERRAKYKREPVSULT 477 | |
| Qy 322 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSLAUV 378 | |
| Db 478 ALLLGLTMTGIAAGVGTGTALVATQFOQLOAAHDDDLKEVEKSIITNLEKSLTSLSEV 537 | |
| Qy 379 VLQNRRLDLLTAERGCTCLFLGEECCYVYVNSQGIIVTEKVEIRDRIQRAABELRNTGPW 438 | |
| Db 538 VLQNRRLDLLFLKEGGLCAALKEECCFYADHTGLVDRDSMAKRLERLSQKLFESQQGW 597 | |
| Qy 439 --GLLSQMPW-----ILPFLGPAAIILLLFGPCIFNLLVNFVSSRIEAVK 484 | |
| Db 598 FEGLFNK-SPWFTTLSTIMGPLIILLIILLLFGPCILNRLVQIKDRISVVQ 648 | |

RESULT 25

VQVWV2

env polyprotein - Friend mink cell focus-forming virus

N/Contains: knob protein gp70; spike protein p15E

C/Species: Friend mink cell focus-forming virus

C/Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999

C/Accession: A03989

R/Koch, W.; Zimmermann, W.; Oliff, A.; Friedrich, R.

J. Virol. 49, 828-840, 1984

A/Title: Molecular analysis of the envelope gene and long terminal repeat of Friend mink

A/Reference number: A03989; MUID: 84138778; PMID: 6321768

A/Accession: A03989

A/Molecule type: DNA

A/Residues: 1-627 <KOC>

A/Cross-references: GB:M12528; NID:g331918; PID:AAA46483.1; PID:g331920

C/Genetics:

A/Gene: env

C/Superfamily: type C retrovirus env polyprotein

C/Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

F/1-32/Domain: signal sequence #status predicted <SIG>

F/33-431/Product: knob protein gp70 #status predicted <KN3>

F/432-627/Product: spike protein p15E #status predicted <SPK>

F/43,58,288,320,360,364,396/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.1%; Score 406; DB 1; Length 627;

Best Local Similarity 29.3%; Pred. No. 2.2e-22;

Matches 108; Conservative 64; Mismatches 145; Indels 52; Gaps 10;

| | |
|--|--|
| Qy 160 THRLVSLFNTTLTGLHEVAQNPTNCWICLPNFRPYVSIPIVPEQWNNFSTEINTTSVL 219 | |
| Db 272 TGDRLNLVPGAYQALNLTSPDKTQECWCLVSGPPYEGVAVLGTYSNHTSAPNCNSVA 331 | |
| Qy 220 VGPLVSNLEIT-----HTSNLTCVKFSNTYTTNSQCIRWVTPPTQIVCLPSGIF 269 | |
| Db 332 SQHKLTLSEVTGQGLCVGAVPKTHQALC-----NTQNTSD-----GSYLAAPAGTI 379 | |
| Qy 270 FVCGTSAYRCLNGS-----SESMLFLSFLVPPMTIYTEQDLYSVVISKPNKRVPI---LP 322 | |
| Db 380 WACNTGLTPTCLSTTVLNLTTDVCVILWELWPKVTHSPSYVYHQFERRAKYKREPVSULTA 439 | |
| Qy 323 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSLAUV 379 | |
| Db 440 LLLGLTMTGIAAGVGTGTALVATQFOQLOAAHDDDLKEVEKSIITNLEKSLTSLSEV 499 | |
| Qy 380 LQNRRLDLLTAERGCTCLFLGEECCYVYVNSQGIIVTEKVEIRDRIQRAABELRNTGPW- 438 | |

| | |
|--|--|
| Db 500 LQNRRLDLLFLKERGLCAALKEECCFYADHTGLVDRDSMAKRLERLTORQKLFESSQGW 559 | |
| Qy 439 -GLLSQMPW-----ILPFLGPAAIILLLFGPCIFNLLVNFVSSRIEAVK----- 484 | |
| Db 560 EGLFNR-SPWFTTLSTIMGPLIILLIILLLFGPCILNRLVQFKDRISVVQALVLTQQYH 618 | |
| Qy 485 ----LQMEP 489 | |
| Db 619 QLKPLEYEP 627 | |

Search completed: April 19, 2004, 10:34:36

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:33:58 ; Search time 17 Seconds
(without alignments)
1647.867 Million cell updates/sec

Title: US-10-069-883-1
Perfect score: 2871
Sequence: 1 MALPYHFLFTVLLPSFTLT.....PPEISGAQLLRPSAGSS 538

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 627 | 21.8 | 563 | 1 | ENV_BAEVM |
| 2 | 620.5 | 21.6 | 574 | 1 | ENV_SRV2 |
| 3 | 601.5 | 21.0 | 574 | 1 | ENV_SRV2R |
| 4 | 599 | 20.9 | 587 | 1 | ENV_SRV1 |
| 5 | 593.5 | 20.7 | 567 | 1 | ENV_AVISN |
| 6 | 593.5 | 20.7 | 586 | 1 | ENV_MPMV |
| 7 | 549 | 19.1 | 575 | 1 | ENV_SNRVH |
| 8 | 502 | 17.5 | 582 | 1 | ENV_AVIRE |
| 9 | 424 | 14.8 | 640 | 1 | ENV_RMCV |
| 10 | 418.5 | 14.6 | 665 | 1 | ENV_MVMO |
| 11 | 414 | 14.4 | 676 | 1 | ENV_MLVFF |
| 12 | 412.5 | 14.4 | 662 | 1 | ENV_FSVGA |
| 13 | 412 | 14.4 | 675 | 1 | ENV_MLVFS |
| 14 | 411.5 | 14.3 | 662 | 1 | ENV_FVLVB |
| 15 | 410 | 14.3 | 676 | 1 | ENV_MLVFP |
| 16 | 409.5 | 14.3 | 642 | 1 | ENV_FVLGL |
| 17 | 409 | 14.2 | 645 | 1 | ENV_FSVSM |
| 18 | 408 | 14.2 | 668 | 1 | ENV_FLVCG |
| 19 | 406.5 | 14.2 | 636 | 1 | ENV_MCFE |
| 20 | 406.5 | 14.2 | 669 | 1 | ENV_MLVAV |
| 21 | 402 | 14.0 | 640 | 1 | ENV_MCFE3 |
| 22 | 401 | 14.0 | 665 | 1 | ENV_MLVDR |
| 23 | 399.5 | 13.9 | 639 | 1 | ENV_FLVSA |
| 24 | 399.5 | 13.9 | 679 | 1 | ENV_FSVSM |
| 25 | 399 | 13.9 | 665 | 1 | ENV_MLVRC |
| 26 | 389.5 | 13.6 | 667 | 1 | ENV_GALV |
| 27 | 387.5 | 13.5 | 661 | 1 | ENV_MLVCH |
| 28 | 384.5 | 13.4 | 666 | 1 | ENV_MLVBO |
| 29 | 373 | 13.0 | 671 | 1 | ENV_FSNV1 |
| 30 | 330.5 | 11.5 | 211 | 1 | ENV_MSVFB |
| 31 | 327.5 | 11.4 | 201 | 1 | ENV_MLVKT |
| 32 | 213.5 | 7.4 | 486 | 1 | ENV_HLV2 |
| 33 | 206 | 7.2 | 488 | 1 | ENV_HTLIC |

RESULT 1

| | | | | | |
|-----|-------|------|---|------------|--------------------|
| 7.1 | 203 | 488 | 1 | ENV_HTL1M | P23064 human t-cel |
| 7.0 | 201 | 488 | 1 | ENV_HTL1F | Q03817 human t-cel |
| 7.0 | 201 | 488 | 1 | ENV_HTL1N | Q03816 human t-cel |
| 6.5 | 187.5 | 534 | 1 | ENV_HTL1A | P03381 human t-cel |
| 6.1 | 176 | 534 | 1 | ENV_FSVST | P03392 feline sarc |
| 5.8 | 167.5 | 515 | 1 | ENV_BLV | P25057 bovine leuk |
| 5.7 | 163.5 | 515 | 1 | ENV_BLV | P51519 bovine leuk |
| 5.6 | 159.5 | 515 | 1 | ENV_BLVBS | P25507 bovine leuk |
| 5.5 | 157.5 | 515 | 1 | ENV_BLVAF | P25504 bovine leuk |
| 5.5 | 157.5 | 515 | 1 | ENV_BLVAV | P25505 bovine leuk |
| 5.3 | 153.5 | 515 | 1 | ENV_EBOIC | P03380 bovine leuk |
| 5.3 | 152 | 676 | 1 | VGP_EBOIC | Q66810 ebola virus |
| 5.2 | 149.5 | 515 | 1 | ENV_BLV2 | P25506 bovine leuk |
| 5.2 | 148.5 | 603 | 1 | ENV_RSVR | P03397 rous sarcom |
| 5.0 | 143 | 676 | 1 | VGP_EBOSM | Q66798 ebola virus |
| 4.9 | 141 | 676 | 1 | VGP_EBOSB | Q66814 ebola virus |
| 4.9 | 141 | 681 | 1 | VGP_MABVM | P32553 marburg vir |
| 4.9 | 139 | 681 | 1 | VGP_MABVP | P32554 marburg vir |
| 4.8 | 139 | 681 | 1 | VGP_MABVP | P32554 marburg vir |
| 4.5 | 130 | 257 | 1 | ENV_AVIR3 | P33498 avian retro |
| 4.5 | 128 | 677 | 1 | VGP_EBORS | Q89853 ebola virus |
| 4.5 | 128 | 677 | 1 | VGP_EBORS | Q89853 ebola virus |
| 4.4 | 126.5 | 246 | 1 | ENV_RSVSR | P03397 rous sarcom |
| 4.4 | 124 | 174 | 1 | ENV_AVISU | P03398 avian sarco |
| 4.3 | 123 | 677 | 1 | VGP_EBORS | Q66799 ebola virus |
| 4.3 | 122 | 436 | 1 | ENV_FLVCS | Q02077 feline leuk |
| 4.1 | 119 | 676 | 1 | VGP_EBOG4 | O11457 ebola virus |
| 4.0 | 116 | 77 | 1 | ENV_SMSAV | P03384 simian sarc |
| 4.0 | 115.5 | 408 | 1 | ENV_RSFFV | P03389 rauscher sp |
| 3.9 | 113 | 410 | 1 | ENV_FLVCA | Q02076 feline leuk |
| 3.8 | 109.5 | 551 | 1 | Y900_METJA | Q58310 methanococ |
| 3.8 | 108.5 | 855 | 1 | ENV_FIVU8 | Q04995 feline immu |
| 3.7 | 107 | 944 | 1 | VGLB_HSVT2 | Q9W15 herpesvirus |
| 3.7 | 106.5 | 854 | 1 | ENV_FIVMO | Q05312 feline immu |
| 3.7 | 106 | 676 | 1 | VGP_EBOEC | P87671 ebola virus |
| 3.7 | 105.5 | 676 | 1 | VGP_EBOZ5 | P87666 ebola virus |
| 3.7 | 105.5 | 676 | 1 | VGP_EBOZM | Q05320 ebola virus |
| 3.6 | 104.5 | 855 | 1 | ENV_FIVU2 | Q04993 feline immu |
| 3.6 | 104 | 768 | 1 | ENV_SIV1 | P27757 simian immu |
| 3.6 | 104 | 854 | 1 | ENV_SIVCZ | P17281 chimpanzee |
| 3.5 | 101.5 | 990 | 1 | ENV_OMVVS | P16899 ovine lenti |
| 3.5 | 101.5 | 5703 | 1 | MUSB_HUMAN | Q9H884 homo sapien |
| 3.5 | 101 | 445 | 1 | ENV_HLVFR | P03395 friend muri |
| 3.5 | 101 | 860 | 1 | ENV_HV2BE | P18094 human immu |
| 3.5 | 101 | 3005 | 1 | ZFH2_DROME | P28167 drosophila |
| 3.5 | 99.5 | 866 | 1 | ENV_GAEVC | P31626 caprine art |
| 3.4 | 98.5 | 856 | 1 | ENV_FIVPE | P16090 feline immu |
| 3.4 | 98.5 | 2517 | 1 | NCR2_HUMAN | Q9Y618 h nuclear r |
| 3.3 | 96 | 1239 | 1 | NME3_MOUSE | Q01098 mus musculu |
| 3.3 | 95.5 | 856 | 1 | ENV_FIVU1 | P29995 rattus norv |
| 3.3 | 95.5 | 2701 | 1 | IP3S_RAT | Q03804 feline immu |
| 3.3 | 95 | 810 | 1 | LRC8_MOUSE | Q80W95 mus musculu |
| 3.3 | 95 | 1148 | 1 | VGLM_PUMMH | Q21400 pumala vir |
| 3.3 | 95 | 1148 | 1 | VGLM_PUMMK | P41265 pumala vir |
| 3.3 | 94 | 1148 | 1 | VGLM_PUMMP | P41266 pumala vir |
| 3.3 | 94 | 670 | 1 | VGLM_SCHPO | O13854 schizosacch |
| 3.3 | 94 | 810 | 1 | LRC8_HUMAN | Q81W6 homo sapien |
| 3.3 | 93.5 | 409 | 1 | ENV1_FRSFV | P03393 friend sple |
| 3.3 | 93.5 | 1328 | 1 | FINC_PLEWA | Q91289 pleurodeles |
| 3.3 | 93.5 | 1333 | 1 | UTX_MOUSE | Q70546 mus musculu |
| 3.3 | 93 | 1333 | 1 | HENA_IHAK6 | P19589 influenza a |
| 3.2 | 93 | 565 | 1 | AGAI_YEAST | P33233 saccharomyc |
| 3.2 | 93 | 725 | 1 | ENV2_FRSFV | P03394 friend sple |
| 3.2 | 92.5 | 409 | 1 | ENV2_FRSFV | P16996 influenza a |
| 3.2 | 92.5 | 565 | 1 | HENA_IHAK7 | P17001 influenza a |
| 3.2 | 92.5 | 565 | 1 | HENA_IHAK7 | P17001 influenza a |
| 3.2 | 92 | 470 | 1 | LEU2_NITEU | Q82W19 nitrosomona |
| 3.2 | 92 | 934 | 1 | PKN1_CHLMU | Q9K92 chlamydia m |
| 3.2 | 92 | 1840 | 1 | M2C1_RAT | P21139 rattus norv |
| 3.2 | 91.5 | 854 | 1 | ENV_SIV1 | Q02837 simian immu |

ALIGNMENTS

```

ENV_BAEVM          STANDARD;          PRT;          563 AA.
ID  ENV_BAEVM
AC  P10269;
DT  01-MAR-1989 (Rel. 10, Created)
DT  01-MAR-1989 (Rel. 10, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  ENV polyprotein precursor (Coat polyprotein) [Contains: Outer membrane
DE  protein GP70; Transmembrane protein P20E].
GN  ENV.
OS  Baboon endogenous virus (strain M7).
OC  Viruses; Retroviridae; Retroviridae; Mammalian type C retroviruses.
OX  NCBI_TaxID=11764;
RN  [1]_
RP  SEQUENCE FROM N.A.
RA  Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;
RT  "The entire nucleotide sequence of baboon endogenous virus DNA: a
RT  chimeric genome structure of murine type C and simian type D
RT  retroviruses."
RL  Jpn. J. Genet. 62:127-137(1987).
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; D10032; BAA00924.1; -;
DR  EMBL; X05470; CAA29028.1; -;
DR  PIR; J02622; VCMVM7.
DR  HSSP; P03385; IMOF.
DR  InterPro; IPR002050; Env polyprotein.
DR  Pfam; PF00429; Env polyprotein; 1.
KM  Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT  SIGNAL 1 20
FT  CHAIN 21 346
FT  TRANSMEMBRANE PROTEIN GP70 (POTENTIAL).
FT  CHAIN 347 563
FT  TRANSMEMBRANE PROTEIN P20E (POTENTIAL).
FT  CARBOHYD 113 113
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 219 219
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 229 229
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 264 264
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 282 282
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 292 292
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 306 306
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 312 312
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 321 321
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 339 339
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
FT  CARBOHYD 469 469
FT  N-LINKED (GLCNAC. .) (POTENTIAL).
SQ  SEQUENCE 563 AA; 61879 MW; 9573137DC4620BB7 CRC64;

Query Match
Best Local Similarity 31.8%; Score 627; DB 1; Length 563;
Matches 158; Conservative 87; Mismatches 169; Indels 86; Gaps 16;

QY 31 SSPYQFLWRMQPQGNIDAPSVRSLSKGTPTTAHTMPRNCYHSATLCMHANTHYW--- 87
DB 83 SGFLQSC-----PCN-----SVQS-----SVHS-----SCYTSYQCRSGNKTYTAT 120
QY 88 -----TGKMINPSPGGLGVTVTCVYPTQTGMSDGGGVQDAREKHVK 130
DB 121 LLKTQTGTSVDQVLGSGTNKLIQSPCNIGKQSGICWSTTAPTHVSDDGGGLDTR---IK 177
QY 131 EVISQLTRVHGTSSP---VKGLDLSKLHETLRTHRLVSLNTLTGLHVSQAQNPNCW 187
DB 178 SVORKLEELKALPELQVHPLAIPKVRNLMVDAQTLINATNALLMSTSLVDDCW 237
QY 188 ICPLNFRPVYSVPPEQWNNFSTEL--NTTSVLVGPLVSNLEITHSTNLTC----- 237
DB 238 LCLKLG--PPTPLAENFLLSVYTRSSDNISCLIIIPLL--VQPMQFNSNCSLSPSYNS 293
QY 238 -----VKFSNTTNTTNSQCIWRTPPTQIVCLPISGIFVCGTS--AYRCLNGSSSMC 288

```

```

DB 294 TEIDILGHVAFNSCTSIYN-----VTGP-----ICAVNGSVFLCGNNMAYTYLPTNWTGLC 344
QY 289 FLSFLVPPMTIYTEQD-----LYSYVSKPRNKRVPILPFGVIGAGVLGATGIGIGITTS 343
DB 345 VLATLLPDIDIIIPGDEFPVPIPAIDHFIYRPK-RAIQFIPLLAGLIGITAAFTTGATGLGVS 403
QY 344 TQFYKLSQELNDEMVERADSLVTLQDLQSLAAVVLQNRRLDILTAERGCTCLFLGEE 403
DB 404 VTQYTKLSNQLISDVQLASTIQDLDQVDSLAENVLQNRGLDILTAERGCTCLALQEK 463
QY 404 CCYVYNSQSGIVTEKVKKEIRDRIOORABEERNTOFGWLLSQWMPWILPFGGLPAAIILL 463
DB 464 CCYVYNSQSGIVTEKVKKEIRDRIOORABEERNTOFGWLLSQWMPWILPFGGLPAAIILL 523
QY 464 FGPCIENLNVFVSSRIEAV 483
DB 524 IGPCIFNRLTAFINDKLNII 543

RESULT 2
ENV_SRV2          STANDARD;          PRT;          574 AA.
ID  ENV_SRV2
AC  P51515;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  ENV polyprotein (Coat polyprotein) [Contains: Coat protein GP70; Coat
DE  protein GP20].
GN  ENV.
OS  Simian retrovirus SRV-2.
OC  Viruses; Retroviridae; Retroviridae; Betaretrovirus.
OX  NCBI_TaxID=39068;
RN  [1]_
RP  SEQUENCE FROM N.A.
RA  Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J.,
RA  Luciw P.A.;
RT  "Sequence relationships of type D retroviruses which cause simian
RT  acquired immunodeficiency syndrome."
RL  Virology 157:317-329(1987).
CC  -I- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; M16605; AAA47563.1; ALT_INIT.
DR  HSSP; P03385; IMOF.
DR  InterPro; IPR002050; Env polyprotein.
DR  Pfam; PF00429; Env polyprotein; 1.
KM  Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT  CHAIN 1 382
FT  COAT PROTEIN GP70.
FT  CHAIN 383 574
FT  COAT PROTEIN GP20.
FT  TRANSMEM 386 406
FT  TRANSMEM 515 535
FT  TRANSMEM 117 117
FT  CARBOHYD 233 233
FT  CARBOHYD 260 260
FT  CARBOHYD 267 267
FT  CARBOHYD 288 288
FT  CARBOHYD 298 298
FT  CARBOHYD 312 312
FT  CARBOHYD 318 318
FT  CARBOHYD 327 327
FT  CARBOHYD 345 345
FT  CARBOHYD 475 475
SQ  SEQUENCE 574 AA; 62952 MW; 662B1E11A437E055 CRC64;

Query Match
21.6%; Score 620.5; DB 1; Length 574;

```


OC Viruses; Retrovirus; Retroviridae; Betaretrovirus.
 OX NCBI_TaxID=11942;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86151668; PubMed=3006247;
 RA Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,
 RA Luciw P.A.;
 RT "Nucleotide sequence of SRV-1, a type D simian acquired immune
 RT deficiency syndrome retrovirus.";
 RL Science 231:1567-1572(1986).
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M1841; AAA47733.1; -;
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; glycoprotein; Polyprotein; Transmembrane; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 395 COAT PROTEIN GP70.
 FT CHAIN 396 587 COAT PROTEIN GP20.
 FT TRANSMEM 528 548 POTENTIAL.
 FT TRANSMEM 528 548 POTENTIAL.
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 587 AA; 64474 MW; 80939DD3BF65A8D CRC64;
 Query Match 20.9%; Score 599; DB 1; Length 587;
 Best Local Similarity 30.2%; Pred. No. 1.2e-39;
 Matches 164; Conservative 86; Mismatches 191; Indels 102; Gaps 20;
 QY 24 PCR---MTSSPYQBFLLMRQPGNIDAPSYRLS-----XGTFPTAHTH--- 67
 DB 41 PDCAGGYSSPTNSLT-----TVSCSTYATSVNLSKWCQCVSTPTTASPTHGSC 93
 QY 68 -----MPRNCVHSATLCHWANTHWYTKMI-----NPS----- 95
 DB 94 PSCNSQSYDSVHATCNHYQCCTGNKYLATMIRDSKPSGGDNVTILGNQNLI 153
 QY 96 --CP--GGLGVTVCTYFTQTGMDGGQVQDQAREKHEKVISQLTRVHGTSSPYKGLDLS 152
 DB 154 AGCPENKKGQVVCWNSQPSVHMSDGGQPDQKREIIVNKKFELHKSFLPELSYHPLALP 213
 QY 153 KLIH--ETLRTHRLVLFN--TTLTGLHVEISAQ--PTNCWICLPLNPPYVSIPIVPEQW 206
 DB 214 EARGKEKIDAH-----FDLATAVHSLNVSORQAEBCWLC-----RSGDPVPLALP 264
 QY 207 NPFSTEINT-----TSVLVGP--LVSNLEIHTSNLCV--KFSNTTYT-----T 247
 DB 265 DNTSCNSTFFNCNSCLITPFLVQPFNFTHS---VCLYADYQNNSDIDVGLAGFT 321
 QY 248 NSQCIRWTPPTQIVCLPSGIFTCVGTG--AYRCLNGSSSMCLSLFLVPMIYEQD-- 304
 DB 322 NCSSYINISKPSPLCAPNSVFCVGNKKAITYLPTNWTGSCVLTALLDIDIPGSEPV 381

QY 305 ---LYSVISKPRMKRVPIILFFVIGAGVLCALGTGGITTTQFYKLSQELNGDMERV 361
 DB 382 PIPALDHLGRPK-RAQFIPLVIGLITTAVSTAGTGLVSLTQVTKLSHQLISDVQAI 440
 QY 362 ADSLVITLQDQLNSLAADVVLONRALDLTAERGCTCLFLGEECCYYVNSGSIIVTEKVEI 421
 DB 441 SSTIQDQLQDQVDSLAEVVLONRRGLDLTAQGGICLALQEKCCFYANKSGIIVRDKINL 500
 QY 422 RDRIQRAEELRNTGPMGLLSOMPMILPFLGLAAIILLLFGPCIFNLLVNVSSRIE 481
 DB 501 QDDLEKRRKQLIDNPFWTGPHGLLPYVWPLGLLCLLLVLSFGPIFNKLMTFIKHIE 560
 QY 482 AVK 484
 DB 561 SIQ 563
 RESULT 5
 ENV_AVISN STANDARD; PRT; 567 AA.
 AC P31796;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein (Coat polyprotein) [Contains: Coat protein GP73; Coat
 DE protein GP22].
 DE ENV.
 GN ENV.
 OS Avian spleen necrosis virus.
 OC Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.
 NCBI_TaxID=11899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Clone PPB101;
 RX MEDLINE=92219390; PubMed=1313915;
 RA Kewalramani V.N., Fanganiban A.T., Emerman M.
 RT "Spleen necrosis virus, an avian immunosuppressive retrovirus, shares
 RT a receptor with the type D simian retroviruses.";
 RL J. Virol. 66:3026-3031(1992).
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M87666; -; NOT_ANNOTATED_CDS.
 DR FIR; A38212; VCFVAS.
 DR HSP; P03385; IMOF.
 DR InterPro; IPR008965; Cellul bind.
 DR InterPro; IPR002050; Env_polyprotein.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; glycoprotein; Transmembrane; Polyprotein.
 FT CHAIN 1 397
 FT CHAIN 398 567 COAT PROTEIN GP73.
 FT CHAIN 567 567 COAT PROTEIN GP22.
 FT TRANSMEM 402 418 POTENTIAL.
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 274 274 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 319 319 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 335 335 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 491 491 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 567 AA; 61596 MW; 204EA57C32159175 CRC64;
 Query Match 20.7%; Score 593.5; DB 1; Length 567;
 Best Local Similarity 33.58%; Pred. No. 3.2e-39;
 Matches 154; Conservative 75; Mismatches 176; Indels 55; Gaps 13;

QY 66 THMPNCHSATLCHNANTHWY-----GKMINSPCGGLGVTCWYFT 110
Db 121 TQMSHCTEYKQECTLLGKTYTALQTKLGSVDGKGLQASCTGTGKPCWDPVA 180
QY 111 QTMSDGGGVDQAREKHVEVISQTRVHGTSSPYKGLDLSKLHETLRTLRLVSLFNT 170
Db 181 PVTSDGGGPTDMIREESVRELEIRHSYPSVQYHFLALPR-----SRGVLDLPQ 232
QY 171 T---LTGLHEV-SAQNP---TNCWICLPLNFRPVISIPVQWNNFSTEINTTSVLGVL 223
Db 233 TSDILEATHQVLNATNPKLAENCWLCMTLGTPIPAAI-----TNGNVTLGDNCSLSLPF 287
QY 224 VSN-----LEIT-----HTSNLCVPSNTYTTNSOCIRWTPPTQ-----IVCLPSGIF 269
Db 288 GCNPGSIVDSYAGEADNRIGIPGVYHF-TNCTSIQEVNTESQGNLRLCPLPPGHV 346
QY 270 FVCGTS-AYRCLNGSSSFCPLFLVPMWYTTOD-----LYSVISKPRNKR-VPIPLPV 324
Db 347 FVCGNNMAYTALPNKWIGLCILASIVPDISISGEEP:PLPSIBYTAARRKRAVQFIFLL 406
QY 325 IGAGVLGALGTGIGGITTSTOFFYKLSOELANGMERVADSLVTIQDOLNSLAADVLRNR 384
Db 407 VGLGISGATLAGTGLGVSVHYTHKLSNQLIEDVQALSGETINDIQDLSLAEVLRNR 466
QY 385 ALDLTAERGTCPLGSECCYVNSQGIIVTEKVEIKRDIRRABEELNRTGPGWLLSQW 444
Db 467 GLDLTASQGGTCLALQKCCFYANKSGIVRDKIRKQEDLIERKRALYDNPLWSGLNGF 526
QY 445 MPWILPFGPLAAIILLGLGPCIFNLAVNVSRIEAVK 484
Db 527 LPYLLPLGLPLFGLILFTLGPCLMKTLRIHDKIQAVK 566
RESULT 6
ENV_MPMV
ID ENV_MPMV STANDARD; PRT; 586 AA.
AC P01575;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE GP70; Coat protein GP20].
GN ENV
OS Simian Mason-Pfizer virus (MPMV).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11855;
RN [1]
RL Cell 45:375-385(1986).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12349; AAA4712.1; -
DR PIR; D25839; VCLJMP.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 394 COAT PROTEIN GP70.
FT CHAIN 395 586 COAT PROTEIN GP20.
FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 586 AA; 63882 MW; 3CE7A399D9E2F450 CRC64;
Query Match 20.7%; Score 593.5; DB 1; Length 586;
Best Local Similarity 28.6%; Pred. No. 3.3e-39;
Matches 156; Conservative 89; Mismatches 192; Indels 109; Gaps 16;
QY 24 PKCMTSSSPYOELWRMORPNIDAPYSRLSK-----GTPPTT 63
Db 41 PCDC-----AGSYVSPPINSLTSCSTHTAYSVNLSKWCQVSTPTTP 85
QY 64 AHTH-----MPNCHSATLCHNANTHWYTG----- 89
Db 86 SNTHGSCPGECNTISYDSVHASCYNHYOQCNKTYLTATITGRTPAIGDGNVPTVL 145
QY 90 ---KMINSPCGG-LGVTVWYFTQTGMSDGGVQDOAREKHVEVISQTRVHGTSS 144
Db 146 GTSHNLITAGCPNGKKGQVVCWNSRPSVHISDGGGPDKARDIIVNKKFEELHSLFPEL 205
QY 145 PYKGLDLSKLH-ETLRTHT-----RLVSLFNTLTGLHEVSAQNPTNCWICLPLNFRPY 197
Db 206 SVHPLALPEARGKEKIDATLIDLLATVHSLNASQPSLAE-----DCWLCLQSDGDPV 258
QY 198 VSIPEPEQ-WNNFSTEINTTSVLGP-LVSNLEITHSNLCVKSNTTYTN-----S 249
Db 259 LALPYNDTLCNSFACLSNHSCLTPPLVOPNFNT-DSNCLYAHYQNNSFIDVGLASFT 317
QY 250 QC-----IRWTPPTQIVCLPSGIPVCGTS-AYRCLNGSSSFCPLFLVPMWYTTQ 303
Db 318 NCSSYVNTASKPSNSLCAPSSVFCNNKAYTLPNTWGTSCVATLPLDIDIDIPS 377
QY 304 D-----LYSVISKPRNKEVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDM 358
Db 378 EPVPIPAIDHFLGKAK-RAIQILPLFVGLGITAVSTGAAGLVGSITQYTKLSHQLISDV 436
QY 359 ERVADSLVTLQDOLNSLAADVONRRALDLTLAERGTCPLFLGSECCYVNSQGIIVTEKV 418
Db 437 QAISSITQIDLOQVDSLAELVQLNRGDLTLAEOGGICLALQKCCFYANKSGIVRDKI 496
QY 419 KEIRDIRRABEELNRTGPGWLLSQWMPWILPFGPLAAIILLGLGPCIFNLVNFVSS 478
Db 497 KLODDLERRRQLIDNPFTSFHGFPLVMPPLGLLCLLLVLSGPIIFNKMTFIKH 556
QY 479 RIEAVK 484
Db 557 QIESIQ 562
RESULT 7
ENV_SMRVH
ID ENV_SMRVH STANDARD; PRT; 575 AA.
AC P21412;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Outer membrane
DE protein; Coat protein GP20].
GN ENV
OS Squirrel monkey retrovirus (SMRV-H) (SMRV-HLB).
OC Viruses; Retrovirdae; Retroviridae; Betaretrovirus.
OX NCBI_TaxID=11856;
RN [1]
RL SEQUENCE FROM N.A.
RX MEDLINE=89073750; PubMed=3201749;

RA Oda T., Ikeda S., Watanabe S., Hataushika M., Akiyama K.,
RA Mitsuobu F.;
RT "Molecular cloning, complete nucleotide sequence, and gene structure
of the provirus genome of a retrovirus produced in a human
lymphoblastoid cell line.";
RL Virology 167:468-476(1988).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC
EMBL; M23385; AAA66454.1; -;
EMBL; M23385; AAA66455.1; -;
PIR; D34827; VCLTHD.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 17
FT CHAIN 18 386 OUTER MEMBRANE PROTEIN.
FT CHAIN 387 575 COAT PROTEIN GP20.
FT TRANSMEM 387 403 POTENTIAL.
FT TRANSMEM 518 534 POTENTIAL.
FT SITE 447 481 IMMUNOSUPPRESSIVE PEPTIDE.
FT CARBOHYD 126 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 139 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 575 AA; 62245 MW; 7210DECC53936669 CRC64;
Query Match 19.1%; Score 549; DB 1; Length 575;
Best Local Similarity 30.6%; Pred. No. 1.1e-35;
Matches 148; Conservative 83; Mismatches 148; Indels 105; Gaps 18;
QY 66 TEMPRNCHGATLCMHANTHYMTGK- - - - -INP- - - - -SQPGGLGVTV 104
DB 109 SQHSSCYSSFSQCTGQNNYFTAILQRYKSTETNPVTSGLQPHGVLAGDGVGKSV 168
QY 105 CWTPYQTGMSDGGVQDQAREKHVEVI- - - - -SOLTRVH- - - - -GTSPYKGLDLSKLH 155
DB 169 CWNQAPIHVS DGGGPDQAVRELYVQKIELVTSQSPKLSYHPLARSKP-RGPDID- - - 224
QY 156 ETLRTHRLVSLFWTLTGLHEVSAQNPTNCWICL- - - - -PLNF- - - - - 194
DB 225 - - - - -AQMIDIUSATHQALNISNPSLAQNCWLCNLQGTSMPLAFPNVNSSFNASQNNCT 278
QY 195 - - - - -RPYVSIPIPEQ-WNNF- - - - -STINTTSLVGLVSLNLEITHSTNLTCVKFNNTVTTNS 249
DB 279 PSLPFRVQWPSPQVYPCFFKGAQNGSFDIPVG-VAN- - - - -FVNCSSSNH 323
QY 250 QCTRWTPPTQIVCLPSGIFPVCGTS-AYKCLNGSSSMCFSLFLYPMTIYEQD- - - - 304
DB 324 - - - - -SEALCPGPGQAFVCGNNLAFALPANWTGSCVLAALLPDIDISGDDPVP 374
QY 305 -LYSVISIKPRNKRVPILPVIAGVILGALGTGIGGITTSQFYKLSQELNGDMERVAD 363
DB 375 PTFDYIAGR-QKRAVTLIPLLVGLGVSTAVATAGLGVAVQYTKLSHQLINDVQALSS 433
QY 364 SLVTLPQQLNSLAIVLQNRALDLITAEKGGTCLFLBEGCCYVYVQSGIVTEKVEIRD 423
DB 434 TINDLPQQLSLAEVILQNRGLDLITAEQGGICLALQERCCFYANKSGIVRDKIKNLOE 493
QY 424 RIQR- - - - -AEEELNTGPEWLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLIVNPFSSR 479

DB 494 DLEKRRKALADNLF LTGLINGLL- - - - -PYLLPFLGLPFAILFFSFAPWILLRVVTAIRQ 549
QY 480 IEAV 483
DB 550 INSL 553
RESULT 8
ENV_AVIRE STANDARD; PRT; 582 AA.
AC P03399; 1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein (Coat polyprotein) [Contains: Coat protein GP73; Coat
protein GP22].
GN ENV.
OS Avian reticuloendotheliosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A;
RX MEDLINE=85009850; PubMed=6090694;
RA Wilhelmsson K.C., Eggleston K., Temin H.M.;
RT "Nucleic acid sequences of the oncogene v-rel in
reticuloendotheliosis virus strain T and its cellular homolog, the
proto-oncogene c-rel.";
RL J. Virol. 52:172-182(1984).
CC -I- PWM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -I- MISCELLANEOUS: STRAIN A IS A HELPER VIRUS OF THE STRAIN T.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC
EMBL; K02537; AAA99198.1; -;
EMBL; X01455; CAA25686.1; -;
PIR; A03999; VCVDFAR.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Transmembrane; Polyprotein.
FT CHAIN 1 391
FT CHAIN 392 582 COAT PROTEIN GP73.
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 582 AA; 64138 MW; CD2560ADF026D32 CRC64;
Query Match 17.5%; Score 502; DB 1; Length 582;
Best Local Similarity 28.6%; Pred. No. 5.7e-32;
Matches 159; Conservative 75; Mismatches 192; Indels 130; Gaps 19;
QY 13 LLPGFTLTAPPCCRCMTSSSPYQEFLEWMPORPGNIDAFYSRLSKGTPTFTTAHTMPRNC 72
DB 101 IIPS-VQEQPCPCCLTIA- - - - -MHAN-THYTGKMINPSCPGGLGTV 104
QY 73 YHSATLC- - - - -YKAEQECTLLGKTYFTAILQKTLGSDYEDGNKLLQASCTGIWETSLMGPSCP 180
DB 126 YKAEQECTLLGKTYFTAILQKTLGSDYEDGNKLLQASCTGIWETSLMGPSCP 180
QY 105 CWTYFTQTGMSDGGVQDQAREKHVEVISOLTRVHGTSSPYKGLDLSKLHETLRTHRL 164
DB 181 C- - - - -LDGGGPTDRF-GRICAGLEBIIRHSVSYQHFLALPR- - - - -PRG 222
QY 165 VSLFNTT- - - - -LTGLHEV-SAQN- - - - -TNCWICLPLNFRPYVSIIPVQWNNFSTEINTTS 217

Db 223 VLDPOTSDILEATHQVLTATNPQLAENCLWLTGL-----TSPQPSRRMAMS-----LS 273
 QY 218 VLVGVLVSNLEITHSSN-----LTC-----VKFSNTYTTNSQCIRWVTPPTQI 261
 Db 274 MEIIVLASLSGATHRVNRQCLLCREADNRTGIPGVVHFTNCTSIQESLIRVIVEILRD 333
 QY 262 VCLPSGIFPVCTGSAYRCLNGSSKMCFLSFLVPMWTVTEQD---LYSYVSKRNRK- 317
 Db 334 YVLRVWYLCVQEHAYTALPNKWIGLICILASIVPDMSSIPGEEPILPSIEYTAGRHKRA 393
 QY 318 VPLPPIVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
 Db 394 VQFIPLLVGLGITGATLAGGTGLGVSVHYKLSNQLIEDVQALSGTINDLQDQIDSLAE 453
 QY 378 VVQNRRALDLTAERGCTCLFGBECCEYVYVNSQSGIVTEKVKKEIDRRQRRAEELRNTGP 437
 Db 454 VVQNRRGLDLTAEQGGICLALQKCCFYANKSGIVRKIKQLQEDLLARKRALYDNLPL 513
 QY 438 WGLLSQWMPWLPFLGFLAAIILLFPGCIENLNVFVSSRIEAVK-LQMEPKQSKTK 496
 Db 514 WNLNGFLPYLLPSLGLFLGLLFLTLGPCIKTLTRIHLKIQSKNPRISPAVQATP- 572
 QY 497 IYRPLDRPASPRSDV 512
 Db 573 -----NRDGYPRSMV 582

RESULT 9
 ENV_RMCVFV STANDARD; PRT; 640 AA.
 AC P06445; Q85629; Q89529;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polyprotein) [Contains: Knob protein
 GP70; Coat protein P12E; Coat protein P15E].
 GN ENV.

OS Rauscher mink cell focus-inducing virus.
 OC Viruses; Retroid viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11784;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85237696; PubMed=4009793;
 RA Vogt M., Haggblom C., Swift S., Haas M.;
 RT "Envelope gene and long terminal repeat determine the different
 RT biological properties of Rauscher, Friend, and Moloney mink cell
 RT focus-inducing viruses.";
 RL J. Virol. 55:184-192(1985).
 CC -I- FTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M10100; AAA46528.1; -
 DR PIR; A03390; VCMVRV.
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env_polyprotein.
 DR InterPro; IPR008981; F_MuLV_bind.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 443 KNOB PROTEIN GP70.
 FT CHAIN 444 623 COAT PROTEIN P12E.
 FT CHAIN 444 640 COAT PROTEIN P15E.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 640 AA; 70071 MW; 1E4450343643D799 CRC64;
 Query Match 14.8%; Score 424; DB 1; Length 640;
 Best Local Similarity 30.0%; Pred. No. 9.7e-26;
 Matches 111; Conservative 63; Mismatches 144; Indels 52; Gaps 10;
 QY 160 THRLVSLFNTTTLGLHEVSAQNPTNCWICPLNFRPVYISIPVEQWNNSTEINTSVL 219
 Db 284 TGDRLNLVDGAYQALNLTSPDKTQECWCLVAEPYEGVAVLGTYSNHTSAPTNCVA 343
 QY 220 VGPLVSNLEIT-----HTSNLTGVKFSNTTNTNSQCIRWVTPPTQIVCLPSGIF 269
 Db 344 SQHKLITLSEVTVGQGLCIGTVPKTHQALC-----NTLTKNK-----GSYLVAAGTT 391
 QY 270 FVQTSAYRCLNGS-----SESMCFSLFVPPMTIYEQDLYSYVISKPRNKRPVI---LP 322
 Db 392 WACNTGLTCLSATVNLRTTDDYCVLVELMWRVTHPPSYVYVYQFESYRHKRPFVSLTLA 451
 QY 323 FVIGAGVLGALGTGIGGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSLAAV 379
 Db 452 LLLGLTGGIAGVGITGTTALVATQFQQLHAAVQDDLEKEVKSITNLEKSLTSLSEV 511
 QY 380 LQNRALDLTAERGCTCLFGBECCEYVYVNSQSGIVTEKVKKEIDRRQRRAEELRNTGP- 438
 Db 512 LQNRGLDLLFLKEGGLCAALKBECCFYADHTGLVRDSMAKLRERLTQROKLFESSQGW 571
 QY 439 -GLLSQWMPW-----ILPFLGLPLAAIILLFPGCIENLNVFVSSRIEAVK----- 484
 Db 572 EGFNLR-SPWFTTLLISTIMTGLILLILLFPGCILNRLVQFVKDRISVQVALVLTQOYH 630
 QY 485 ----LQMEPK 490
 Db 631 QLKPLEVEPQ 640

RESULT 10
 ENV_MLVMO STANDARD; PRT; 665 AA.
 ID AC P03385;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ENV polypeptide precursor (Coat polyprotein) [Contains: Knob protein
 GP70; Spike protein P15E; R protein].
 GN ENV.
 OS Moloney murine leukemia virus.
 OC Viruses; Retroid viruses; Retroviridae; Gamma retrovirus.
 OX NCBI_TaxID=11801;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE PMLV-1).
 RX MEDLINE=82035843; PubMed=6169994;
 RA Shinnick T.M., Lerner R.A., Sutcliffe J.G.;
 RT "Nucleotide sequence of Moloney murine leukaemia virus.";
 RL Nature 293:543-548(1981).
 RN [2]
 RP SEQUENCE OF 496-665 FROM N.A. (PROVIRUS).
 RX MEDLINE=81052384; PubMed=6159543;
 RA Sutcliffe J.G., Shinnick T.M., Green N., Liu F.-T., Niman H.L.,
 RA Lerner R.A.;
 RT "Chemical synthesis of a polypeptide predicted from nucleotide
 RT sequence allows detection of a new retroviral gene product.";
 RL Nature 287:801-805(1980).
 RN [3]
 RP SEQUENCE OF 484-665 FROM N.A. (PROVIRUS; CLONE PMLV-201).
 RX MEDLINE=81013872; PubMed=6251454;
 RA Sutcliffe J.G., Shinnick T.M., Verma I.M., Lerner R.A.;
 RT "Nucleotide sequence of Moloney leukemia virus: 3' end reveals
 RT details of replications, analogy to bacterial transposons, and an
 RT unexpected gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).

[4] SEQUENCE OF 470-489 AND 598-665.
 RX MEDLINE=82082389; PubMed=6947213;
 RA Green N., Shinnick T.M., Witte O., Ponticelli A., Sutcliffe J.G.,
 RA Lerner R.A.,
 RT "Sequence-specific antibodies show that maturation of Moloney
 RT leukemia virus envelope polypeptide involves removal of a
 RT COOH-terminal peptide".
 RL Proc. Natl. Acad. Sci. U.S.A. 78:6023-6027(1981).
 RN [5]
 RX X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 514-567.
 RX MEDLINE=96196884; PubMed=8612078;
 RA Fass D., Harrison S.C., Kim P.S.,
 RA "Retrovirus envelope domain at 1.7-A resolution".
 RL Nat. Struct. Biol. 3:465-469(1996).
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J02255; AAS59943.1; -
 CC PIR; A93265; VCVWEM.
 CC PDB; 1MOF; 14-OCT-96.
 CC InterPro; IPR002050; Env polyprotein.
 CC InterPro; IPR008981; F MuLV bind.
 CC Pfam; PF00429; Env polyprotein; 1.
 CC Glycoprotein; Coat protein; Transmembrane; Signal;
 CC 3D-structure.
 CC SIGNAL 1 33
 CC CHAIN 34 469
 CC CHAIN 470 649
 CC CHAIN 650 665
 CC TRANSMEM 610 627
 CC DISULFID 555 562
 CC CARBOHYD 45 45
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 199 199
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 326 326
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 358 358
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 365 365
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 398 398
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CARBOHYD 434 434
 CC N-LINKED (GLCNAC. .) (POTENTIAL).
 CC CONFLICT 655 655
 CC CONFLICT 663 663
 CC Y -> F (IN REF. 3).
 CC Y -> C (IN REF. 3 AND 4).
 CC 516 547
 CC HELIX 548
 CC TURN 549
 CC TURN 552
 CC HELIX 554
 CC TURN 558
 CC TURN 559
 CC SEQUENCE 665 AA; 73301 MW; 12EBA09C8FB93FE2 CRC64;
 Query Match 14.6%; Score 418.5; DB 1; Length 665;
 Best Local Similarity 30.2%; Pred. No. 2.8e-25;
 Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;
 QY 160 THRIVLSFNTLTGLHEVSAQNPNCWICLFLNFRPYVPIPVPEQNNPSTINTSVL 219
 DB 310 TENRLNLVDGAYQALNLTSPDKTQECWCLVAGPPYEGVAVLGTYSNHTSAPNCVSA 369
 QY 220 VGPLYSNLEIT-----HTSNLTCKVPSNTTNTTNSQCIRWTPPTQICVLPSGIF 269
 DB 370 SQHKLTLSVGTGQGLCAVFKTHQALC-----NTQTSSR-----GSYLVAFTGM 417
 QY 270 FVCGTSAVRLNGS-----SESMCFSLFVLPMTTYTQDLYSVYVSKPRNKRVPV---IP 322
 DB 418 WACSTGLPTCISTITLNLTDYCVLVELWPRVTVYHSPSYVYGFFERSNRHKKFVSLTAA 477
 QY 323 FVIGAGVIGAGTGTGGTGT---STOFYKYSORLNGDMERVADSLVTLDOALNSLAAYV 379

SQ SEQUENCE 676 AA; 74049 MW; E18C44511145188B CRC64;
Query Match 14.4%; Score 414; DB 1; Length 676;
Best Local Similarity 29.3%; Pred. No. 6.4e-25;
Matches 110; Conservative 56; Mismatches 147; Indels 62; Gaps 11;
QY 160 THRLVSLFNTTLTGLHEVSAQNPTNCWICLPLNFRPYVSPVPEQNNFSTFNTTSLV 219
DB 320 TGRLLNLVQAYQALNLTNDPKTQECWCLVSGPPYVGVAV-----LQYSNHTSA- 372
QY 220 VPLVSNLEITHSTNLTCVKESNTYTNNSOCIRWTPPTQIVC-----LP 265
DB 373 -----PANGSVASQHKLT-----SEVGRGLCIGTVPKTHALCNLTLLKAGKGSYLVAP 423
QY 266 SGIFPVCTGSAVRCCLNGS-----SESMEFLSVLPPTMTYTBQDLYSYVVISKPRNKRVP 320
DB 424 TGTWACNTGLTPCLSATVLRNTTDCVVLWELPRTVHPPSYVVSQPEKSHRHKRPVS 483
QY 321 --LPVIGAGVLGALTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSL 375
DB 484 LTLALLLGLLTGGTAAGVTGTALVATQFQQLHAAVQDQLKEVEKSIITNLEKSLTSL 543
QY 376 AAVLQNRALDLTAERGGTCLFLGEECCYVYVNSGIVTEKVEIRDIORRAEELRNT 435
DB 544 SEVLQNRGLDLLPLKEGGLCAALKKECCPVADHTGLVRDSMAKLRRLSQRQKLPESS 603
QY 436 GPWLLSOM- -PW-----LPFLGLPAAIILLFPGCINLNLVNFVSSRIEAVK----- 484
DB 604 QGW--FEQWFRNSPWFLLTSTINGPLIILLILLFPGCINLNLVQFVKDRISVVOALVL 661
QY 485 -----LQWEPK 490
DB 662 TQYHQLPLEYEPQ 676
RESULT 12
ENV_FSVGA STANDARD; PRT; 662 AA.
AC P03391; P21446; ST
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein
GP70; Spike protein P15E; R protein].
GN ENV
OS Feline sarcoma virus (strain Gardner-Arnstein) (Ga-Fesv) (Gardner-
Arnstein feline leukemia oncovirus B).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11774;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84115095; PubMed=6319767;
RA Nunberg J.H., Williams M.E., Innis M.A.;
RT "Nucleotide sequences of the envelope genes of two isolates of feline
leukemia virus subgroup B."
RL J. Virol. 49:629-632 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83216322; PubMed=6304347;
RA Elder J.H., Mullins J.I.;
RT "Nucleotide sequence of the envelope gene of Gardner-Arnstein feline
leukemia virus B reveals unique sequence homologies with a murine
mink cell focus-forming virus."
RL J. Virol. 46:871-880 (1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=84131936; PubMed=6321156;
RA Wunsch M., Schulz A.S., Koch W., Friedrich R., Hunsmann G.;
RT "Sequence analysis of Gardner-Arnstein feline leukemia virus
envelope gene reveals common structural properties of mammalian
retroviral envelope genes."
RL EMBO J. 2:2239-2246 (1983).
RN [4]

SEQUENCE FROM N.A.
RX MEDLINE=8044502; PubMed=2823466;
RA Guilhot S., Hampe A., D'Auriol L., Galibert F.;
RT "Nucleotide sequence analysis of the LTRs and env genes of SM-Fesv
and GA-Fesv."
RL Virology 161:252-258 (1987).
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@isb-sib.ch).
CC EMBL; K01209; AAA43052.1; -;
DR EMBL; V01172; CAA24497.1; -;
DR EMBL; X00188; CAA25008.1; -;
DR EMBL; M32026; -; NOT_ANNOTATED_CDS.
DR PIR; A03991; VCVWGF.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 465 KNOB PROTEIN GP70.
FT CHAIN 466 645 SPIKE PROTEIN P15E.
FT CHAIN 646 662 R PROTEIN.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 15 15 MISSING (IN REF. 4).
FT CONFLICT 41 41 V -> I (IN REF. 4).
FT CONFLICT 47 47 LVTGK -> VQNTQ (IN REF. 4).
FT CONFLICT 51 56 FPTMYF -> YPTLV (IN REF. 4).
FT CONFLICT 70 75 IIGTNWPSDQEPFG -> LVGDSWEPIVLDPNVKGAR
YSSSK (IN REF. 4).
FT CONFLICT 80 95 DPMRWQRNT -> KTTDRKKQQT (IN REF. 4).
FT CONFLICT 99 110 NRKQ -> PSLPKGTH (IN REF. 4).
FT CONFLICT 120 123 P -> A (IN REF. 4).
FT CONFLICT 127 127 V -> A (IN REF. 4).
FT CONFLICT 134 134 TYWRPT -> AWKPS (IN REF. 4).
FT CONFLICT 143 148 KGVTOGIQSGGCGGCPYDKAVHSSTTGASEGR ->
RGSSQDTSCEK (IN REF. 4).
FT CONFLICT 158 193 T -> A (IN REF. 4).
FT CONFLICT 208 208 S -> M (IN REF. 4).
FT CONFLICT 215 215 S -> T (IN REF. 4).
FT CONFLICT 223 223 S -> T (IN REF. 4).
FT CONFLICT 232 232 M -> S (IN REF. 4).
FT CONFLICT 238 238 IESKVTPHSGNGGTGPTLVNASIAPLSTPVPAS ->
TGSKVATORTPNTESAPRSVPTMG (IN REF. 4).
FT CONFLICT 264 300
SQ SEQUENCE 662 AA; 73149 MW; 1482088D547CF47 CRC64;
Query Match 14.4%; Score 412.5; DB 1; Length 662;
Best Local Similarity 23.2%; Pred. No. 8.2e-25;
Matches 138; Conservative 89; Mismatches 208; Indels 161; Gaps 21;
QY 44 PG-NIDAPSVRSLSKGTPTTATHM-----PRCYHSATLCMHANTHYWTGKMINPS 95
DB 94 PGYCDQPMRWQRNTFFVCFCHANKKCGGPGQGFCAVWCERTGETYV-----RPT 148
QY 96 -----CFGLGVTVCW---TYFTQTMDSGGG----- 119

Db 149 SSWDYITVKKGVTQGIYQCSCGGWCGPCYDKAVHSSTTGASGGRCNPLILQFTQKGRQT 208
Qy 120 -----VQDQ---AREKHVKEVI 133
Db 209 SWDCPKSWGLRLVRSYDPTALRSVRQVMTIIPQAMGNVLVLPQKPPSRQSIESRV 268
Qy 134 SQ-----LTVHGTSSPYKGLDLSKLHETLTHTRVLSLFTNTLTGLHVSQAQN 182
Db 269 TPHSQNGGTGPGTITLVNAGIAPLSTPTVTPASPKRIGTGDRILNLVQGTLYALNATDPNR 328
Qy 183 PTNCWICLPLNFRPYVSIPIVPEQWNPFSTINTSVLGLVSNLEITHTSNLTCTVKFSN 242
Db 329 TKDCLWCLVSRPPYEGIAL-----LGNYSNQT-----PPSCLSIPOH-----KLT 372
Qy 243 TTYTNSQCIRWVPTPTQIVC-----LPSGIFVFCGTSAYRC-----LNGS 283
Db 373 SEVSGQGLCIGTVPKTHQALCNETQOQHTGAHYLAAPNGTYWACNTGLTPCISMAVLNWT 432
Qy 284 SESMCFSLFVPPMTIITEQDLSYVLSKPNKRVPI---LPFVIG-----AGVLGALGTG 336
Db 433 SD-FCVLTIELWFRVTHQPEVYTHFAKARFRREPISLTVALMGLGLTVGGIAAGVG TG 491
Qy 337 IGGITTTQFYKLSQELNGSMERVAADSLVTLODQNSLAAYVLQNRRLDLTLTAERG GT 396
Db 492 TKALITTAQF-RQLQMAHTDIQALEESISALEKSLTSLSEVLQNRRLDILFLOEGGL 550
Qy 397 CLFLGECCECVYVNGSIVTEKVEIRDIORRABELNTPGWLGLSQWM---PW-----IL 449
Db 551 CAALKKECCFYADHTGLVRNMAKRLKROQLQFDSQQG--FEGWFKSPWFTTLLS 608
Qy 450 PFLGFLAAIILLFGFCIFNLLVNFVSSRIEAVKLQMEPKMSKTKYRPRDRP 505
Db 609 SIMGFLIILLIILLFGFCILNRLVQFVKDRISVVQALILTQQYQKIQY--DPRP 662

RESULT 13
ENV_MLVF5 STANDARD; PRT; 675 AA.
AC P03390;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ENV polypeptide precursor [Contains: Knob protein GP70; Spike protein
DE P15E; R protein].
GN ENV.
OS Friend murine leukemia virus (isolate 57) (F-MuLV).
OC Viruses; Retroviridae; Retroviridae; Mammalian type C retroviruses.
OX NCBI_TaxID=11796;
[1]
RP SEQUENCE FROM N.A.
RA Friedrich R.W., Koch W., von Maydell-Livonius U., Schrewe H.,
RA Zimmermann W.,
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
[2]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=84138778; PubMed=6321768;
RA Koch W., Zimmermann W., Oliff A., Friedrich R.W.;
RT "Molecular analysis of the envelope gene and long terminal repeat of
RT Friend mink cell focus-inducing virus: implications for the functions
RT of these sequences";
EL J. Virol. 49:828-840(1984).
[3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 43-270.
RX MEDLINE=97433384; PubMed=9287219;
RA Fass D., Davey R.A., Hanson C.A., Kim P.S., Cunningham J.M.,
RA Berger J.M.;
RT "Structure of a murine leukemia virus receptor-binding glycoprotein
RT at 2.0-A resolution";
RL Science 277:1662-1666(1997).
CC -!- FUNCTION: INVOLVED IN HOST RANGE; VIRUS RECEPTOR BINDING; CELL
CC FUSION.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X02794; CAA26561.1; -;
DR PDB; 1AOL; 15-OCT-97.
DR InterPro; IPR002050; Env polyprotein.
DR InterPro; IPR008381; F_MuLV_bind
DR Pfam; PF00429; ENV_Polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal;
KW 3D-structure; Zinc; Metal-binding.
FT SIGNAL 1 34 BY SIMILARITY.
FT CHAIN 35 479 KNOB PROTEIN P15E.
FT CHAIN 480 659 SPIKE PROTEIN P15E.
FT CHAIN 660 675 R PROTEIN.
FT TRANSMEM 620 636 POTENTIAL.
FT METAL 55 55 ZINC.
FT DISULFID 80 132
FT DISULFID 106 121
FT DISULFID 107 117
FT DISULFID 155 175
FT DISULFID 167 180
FT DISULFID 212 218
FT DISULFID 346 349
FT DISULFID 395 407
FT DISULFID 376 430
FT DISULFID 437 450
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .)
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .)
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .)
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .)
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .)
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .)
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .)
FT STRAND 44 52
FT TURN 54 55
FT STRAND 58 67
FT TURN 68 69
FT STRAND 75 75
FT STRAND 78 78
FT HELIX 79 83
FT TURN 84 85
FT HELIX 88 90
FT TURN 93 94
FT TURN 97 98
FT STRAND 106 106
FT TURN 107 108
FT TURN 110 111
FT TURN 115 116
FT STRAND 118 118
FT HELIX 119 122
FT TURN 126 127
FT HELIX 135 146
FT TURN 148 149
FT STRAND 152 156
FT TURN 158 159
FT TURN 162 163
FT HELIX 164 167
FT HELIX 170 172
FT TURN 173 174
FT TURN 178 179
FT STRAND 183 185
FT TURN 187 188
FT STRAND 197 201
FT HELIX 205 213
FT TURN 214 215
FT STRAND 220 225
FT HELIX 227 231
FT HELIX 234 236

```

FT TURN 237 237
FT STRAND 239 245
FT TURN 248 250
FT STRAND 253 263
FT HELIX 266 268
SQ SEQUENCE 675 AA; A097038E422BB3D3 CRC64;

Query Match 14.4%; Score 412; DB 1; Length 675;
Best Local Similarity 29.8%; Pred. No. 9.3e-25;
Matches 111; Conservative 57; Mismatches 145; Indels 60; Gaps 11;

QY 160 THTRLVSLFNTLTGLHVEYSAQNPNCWICLPLNRPVYSIPVQWNNFSTEINTSVL 219
DB 320 TGRLLNLVQGAQALNLTNPDKTECWLCLVSGPPYEGVAV-----LGTYSNHTSA- 372
QY 220 VGPLVSNLEITHTSNLTVCVFNSTYNTNSQICRWVTPETQIVC-----LP 265
DB 373 -----PANCVASGSHKLT-----SEVTGRGLCIGTPVKTHQALCNTLTKIDKGSYYLVAP 423
QY 266 SGIFFCVCGTSAYRCLNGS-----SESNCFLSFLVPPMTIYTEQDLYSYVSKPRKRVPI- 320
DB 424 TGTWACNTGLTFCUSATVNLRTDYCVLVELWPRVTVHPPSYVYSQPEKSYRHKREPV 483
QY 321 --LPFVIGAGVLGALGTGGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSL 375
DB 484 LTLALLGLGTWGGIAAGVGTGTALVATQFQQLHAAVQDDLKEVKSITNLEKSLTSL 543
QY 376 AAVLQNRALDLTLAERGGTCLFLGEECCYVNGSGIVTEKVKIRDRIRAEELNRT 435
DB 544 SEVLQNRGLDLFLKKEGGLCAALKECCFVADHTGLVRDSMAKRELTLQKQLPSS 603
QY 436 GPW--GLLSQWMPW-----TLPLGLPLAAIILLFGPCIFNLVNFVSRIEAVK----- 484
DB 604 QGWFEGLFNR--SPWFTTLISTNGPLIILLIILLFGPCILNLFQVDRISVQALVIT 662
QY 485 -----LQMEP 489
DB 663 QQVHQLKPLEYEP 675

RESULT 14
ENV_FLVLVB
ID AC ENV_FLVLVB STANDARD; PRT; 662 AA.
AC P11261; Q85515; Q85516; Q85517;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
DE GP70; Spike protein P15E] (Fragment).
GN ENV.
OS Feline leukemia virus (strain B/lambda-B1).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=103916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88036192; PubMed=2444714;
RA Nicolaissen-Strouss K., Kumar H.P.M., Fitting T., Grant C.K.,
RA Elder J.H.;
RT "Natural feline leukemia virus variant escapes neutralization by a
RT monoclonal antibody via an amino acid change outside the antibody-
RT binding epitope".
RL J. Virol. 61:3410-3415 (1987).
CC -/- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03448; AAA43048.1; -

```

```

DR PIR; A27172; VCMVLB.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 465 KNOB PROTEIN GP70.
FT CHAIN 466 662 SPIKE PROTEIN P15E.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON TER 662 662
SQ SEQUENCE 662 AA; 73132 MW; D4AC18E197230575 CRC64;

Query Match 14.3%; Score 411.5; DB 1; Length 662;
Best Local Similarity 27.8%; Pred. No. 9.9e-25;
Matches 110; Conservative 69; Mismatches 164; Indels 53; Gaps 11;

QY 136 LTRVHGTSPPYKGLDLKSLHETLRTHTRLVSLFNTLTGLHVEYSAQNPNCWICLPLNFR 195
DB 282 ITLVNASIPLSTPVPASPCKRIGTGNRLINLVQGITVIALNVTNPNKTKDCWLCVSRPP 341
QY 196 PYVSIPIVPEOWNFSTEINTSVLGPLVSNLEITHTSNLTVCVFNSTYNTNSQICRWV 255
DB 342 YYEGIAV---LGNYSQNTNPSPCLSDPQHKLTISEVSG-----QGSCIGTV 385
QY 256 TPPTQIVC-----LPSGIFFCVCGTSAYRCLNGSSESMCFLSFLVPP 296
DB 386 PKTHQALCKTKQKHGKTHYLAAPSGTYWACNTGLTFCISMAVLNWTSD-FCVLIELWPR 444
QY 297 MTIYTEQDLYSYVSKPRKRVPI---LPFVIG-----AGVLGALGTGIGITSTSTQYK 349
DB 445 VTYHQPEYVYTHFDKTVLRREPISLTVALMGLTGVGGIAAGVGTGKALLETAPQF-GQ 503
QY 350 LSQELNGDMERVADSLVTLQDQNSLAAVLQNRALDLTLAERGGTCLFLGEECCYV 409
DB 504 LQWAMHTDIQALBESISALEKSLTSSEVVLQNRGLDILFLQEGGLCAALKECCFYAD 563
QY 410 QSGIVTEKVKIEIRDRIQRAEELNRTGFWGLLSQW---PW-----LPFIPGLPAAIILL 462
DB 564 HTGLVRDNMAKRLERLQKQQLFDSQQGM--FEGWFKSPWFTTLISSIMGPLILLILL 621
QY 463 LFGPCIFNLVNFVSRIEAVKLMQEPKQSKTKIY 498
DB 622 LFGPCILNLFVQVDRISVQALILTQQYQKIQY 657

RESULT 15
ENV_MLVFP
ID ENV_MLVFP STANDARD; PRT; 676 AA.
AC P28603;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor [Contains: Knob protein GP70; Spike protein
DE P15E; R protein].
GN ENV.
OS Friend murine leukemia virus (isolate PVC-211) (F-MuLV).
OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
OX NCBI_TaxID=11798;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219364; PubMed=1560524;
RA Masuda M., Remington M.P., Hoffman P.M., Ruscetti S.K.;
RT "Molecular characterization of a neuropathogenic and
RT nonerythroleukemogenic variant of Friend murine leukemia virus
RT PVC-211."

```

| | | | | | | | | |
|---------------------------|--|---|-------------------------|----------------------|---------------------|--|--|--|
| QY | 370 | DQINSIAAVLVQNRRAIDLLTAE | RGTCFLFGECCYYNQSGIVTEK | VEIKEDIRQRA | 429 | | | |
| Db | 538 | KSITSLSVYVQNRRLGDLDF | LKFKBGLCAALKECCFYADHT | GLVRD\$MAKLRERTQ | 597 | | | |
| QY | 430 | EELRNTGPV--GILSOMFEP | ----ILPFLGPLAAIILLFG | FCICIFNLLVNFVSSRIE | 483 | | | |
| Db | 598 | KLFESSQGWFEGLFNR | -SPWFTTLSTINGPLIILL | LFGPCILNRLVQFVKDRISV | 656 | | | |
| QY | 484 | K----- | IQMEPK | 490 | | | | |
| Db | 657 | QALVLTQQVHQKPL | EFQ | 676 | | | | |
| RESULT 16 | | | | | | | | |
| ENV_FLVGL | ID | ENV_FLVGL | STANDARD; | PRT; | 642 AA. | | | |
| AC | P08359; | | | | | | | |
| DT | 01-AUG-1988 | (Rel. 08, Created) | | | | | | |
| DT | 01-AUG-1988 | (Rel. 08, Last sequence update) | | | | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last annotation update) | | | | | | |
| DE | ENV polyprotein precursor | (Coat polyprotein) | [Contains: Knob protein | | | | | |
| DE | GP70; Spike protein P15E] | | ENV. | | | | | |
| OS | Feline leukemia virus | (strain A/Glasgow-1). | | | | | | |
| OC | Viruses; Retrovird viruses; | Retroviridae; Gammaretrovirus. | | | | | | |
| OX | NCBI_TaxID=11769; | | | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | | | |
| RA | MEDLINE=86200439; | PubMed=3009890; | | | | | | |
| RX | Stewart M.A., Warnock M., | Wheeler A., Wilkie N., Mullins J.I., | | | | | | |
| RT | Onions D.E., Neil J.C.; | | | | | | | |
| RT | "Nucleotide sequences of a | feline leukemia virus subgroup A envelope | | | | | | |
| RT | gene and long terminal repeat | and evidence for the recombinational | | | | | | |
| RT | origin of subgroup B viruses." | | | | | | | |
| RL | J. Virol. 58:825-834 | (1986). | | | | | | |
| CC | ----- | | | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | | | | |
| CC | ----- | | | | | | | |
| DR | EMBL; M12500; AAA43053.1; | -; | | | | | | |
| DR | HSPSP; P03385; 1MOF. | | | | | | | |
| DR | InterPro; IPRO02050; Env | polyprotein. | | | | | | |
| DR | InterPro; IPRO08981; F | MuLV bind. | | | | | | |
| DR | Pfam; PF00429; Env | polyprotein; 1. | | | | | | |
| KW | Coat protein; glycoprotein; | Polyprotein; Signal. | | | | | | |
| FT | SIGNAL | 1 33 | | | | | | |
| FT | CHAIN | 34 445 | | | | | | |
| FT | CHAIN | 446 642 | | | | | | |
| FT | CARBOHYD | 35 35 | | | | | | |
| FT | CARBOHYD | 43 43 | | | | | | |
| FT | CARBOHYD | 58 58 | | | | | | |
| FT | CARBOHYD | 91 91 | | | | | | |
| FT | CARBOHYD | 267 267 | | | | | | |
| FT | CARBOHYD | 302 302 | | | | | | |
| FT | CARBOHYD | 307 307 | | | | | | |
| FT | CARBOHYD | 334 334 | | | | | | |
| FT | CARBOHYD | 374 374 | | | | | | |
| FT | CARBOHYD | 390 390 | | | | | | |
| FT | CARBOHYD | 410 410 | | | | | | |
| SEQ | SEQUENCE | 642 AA; 71053 MW; 2DEFS9AEFFC245EC | CRC64; | | | | | |
| Query Match | | | | | | | | |
| | | 14.3%; | Score | 409.5; | DB 1; Length 642; | | | |
| Best Local Similarity | | 28.5%; | Pred. No. | 1.4e-24; | | | | |
| Matches 108; Conservative | | 64; | Mismatches | 152; | Indels 55; Gaps 12; | | | |
| QY | 160 | THTRIVLSFNTLITGLHEVS | AQNPTNCWICLPINFRPYVIS | IPVPEQWNPFSTEINTSVL | 219 | | | |
| Db | 286 | TGPRINLVGGTVLANATD | PNKTKDCTCLVSRPPYEGIAL | ---LGNYSNOTNPPSPC | 342 | | | |

QY 220 VGPLVSNLEIHTSNLCVKFESNTYTTNSOCIRWTPPTQIYC-----LP 265
 Db 343 LSTPQHKLTITSEVSG-----QGCICGTVPKTHQALCNKTKQOQHTGAHYLAAP 389
 QY 266 SGIFVCGTSAIRC-----LNGSSECMFLSFLVPPMTIYTEQDLYSVISKPRNKRVP 320
 Db 390 NGTYWACNTGLTPCISMAVLNWTSD-FCVLIELWPRTVTHQPEYVYTHPAKAVRFRPEI 448
 QY 321 ---LPFVIG-----AGVLGALGTGGITTSQFYKLSQELNGDMERVADSLVTLQDQLN 373
 Db 449 SLTVALMLGGTGVGIAAGVGTGKALLETAPF-RQLQAMHTDIQALEBSISALEKSLT 507
 QY 374 SLAAVVLNRRALDILTAERGGTCLFLGEECCYVNSGIVTEKVKIEIRDIORABEELR 433
 Db 508 SLSEVVLNRRGLDILFLOEGLCAALKECCFADHTGLVRDNWAKLERLQKQQLFD 567
 QY 434 NTGPWGLLSQW-----PW-----ILPFLGLPAAIILLFGLPCIFNLLVNFVSSRIEAVKIQ 486
 Db 568 SQQGW--FEQWFKNSPWFITLSSIMGPLLILLFGLPCILNRLVQFVKDRISVVQAL 625
 QY 487 MEPMQSKTKIYRRPLDRP 505
 Db 626 ILTQYQIQIKY--DPDRP 642

RESULT 17
 ENV_FSVSM STANDARD; PRT; 645 AA.
 ID AC P21445;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein precursor [Contains: Coat protein GP70; Coat protein
 DE P155].
 GN ENV.
 OS Feline sarcoma virus (strain SM) (Sm-FesV).
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11779;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88044502; PubMed=2823466;
 RA Guilhot S., Hampe A., D'Auriol L., Galibert F.;
 RT "Nucleotide sequence analysis of the LTRs and env genes of SM-FesV
 and GA-FesV".
 RL Virology 161:252-258 (1987).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M23025; AAA74004.1; -
 DR HSP; P03385; 1MOF.
 DR InterPro; IPR002050; Env polyprotein.
 DR InterPro; IPR008981; F_MuLV bind.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 36
 FT CHAIN 37 448 COAT PROTEIN GP70.
 FT CHAIN 449 645 COAT PROTEIN P155.
 FT CHAIN 646 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 334 334 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 645 AA; 71594 MW; 4E6DB90A00A43B21 CRC64;
 Query Match 14.2%; Score 409; DB 1; Length 645;
 Best Local Similarity 27.8%; Pred. No. 1.5e-24;
 Matches 116; Conservative 72; Mismatches 170; Indels 60; Gaps 14;
 QY 121 QDAREKHVKYVISOQTRVHGTSPPYKGLDLSKLEHETLRLVSLFNTTITGLHEVSA 180
 Db 255 QSQTSK-----VATQKPTNGT-TPRSVAPATMSPKRIGTRDLINLVQGTYLALNATDP 309
 QY 181 QMPTNCWICPLNFRPYVSIYPVEOWNPSTINTSVLGLVSNLEIHTSNLCVKF 240
 Db 310 NKTQCMWLCLVRPPYEGIAI---LGNYSNQTNPSPSCSLSTPQHKLTITSEVSG 360
 QY 241 SNTTYTNSQCIRWTPPTQIYC-----LPSGIFVCGTSAYRC-----LN 281
 Db 361 -----QGLCIGTVPRTHQALCNKTKQOQHTGAHYLAAPNGTYWACNTGLTPCISMAVLN 413
 QY 282 GSSECMFLSFLVPPMTIYTEQDLYSVISKPRNKRVP---LPFVIG-----AGVLGALG 334
 Db 414 WTSD-FCVLIELWPRTVTHQPEYVYTHFDKAVRFRPEISLTVALMGLTGVGIAAGVG 472
 QY 335 TGIGGITTSQFYKLSQELNGDMERVADSLVTLQDQLNLSLAAVVLNRRALDILTAERG 394
 Db 473 TGTKALLETAPF-RQLQAMHTDIQALEBSISALEKSLTSLSEVVLNRRGLDILFLOQ 531
 QY 395 GTCLFLGEECCYVNSGIVTEKVKIEIRDIORABEELNTPGWLLSQW-----PW----- 447
 Db 532 GLCAALKERCCFYADHTGLVRDNWAKLERLQKQQLFDSDQGW--FEQWFKNSPWFITL 589
 QY 448 ILPFLGLPAAIILLFGLPCIFNLLVNFVSSRIEAVKIQMEPMQSKTKIYRRPLDRP 505
 Db 590 ISSIMGPLLILLFGLPCILNRLVQFVKDRISVVQALI--LTQYQIQIQYDPDRP 645

RESULT 18
 ENV_FLVCG STANDARD; PRT; 668 AA.
 ID AC P21443;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein precursor [Contains: Coat protein GP70; Coat protein
 DE P155].
 GN ENV.
 OS Feline leukemia provirus (clone CFE-6).
 OC Viruses; Retroviral viruses; Retroviridae; Gamma retrovirus.
 OX NCBI_TaxID=11922;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89199802; PubMed=2539525;
 RA Kumar D.V., Berry B.T., Roy-Burman P.;
 RT "Nucleotide sequence and distinctive characteristics of the env gene
 of endogenous feline leukemia provirus".
 RL J. Virol. 63:2379-2384 (1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M25425; AAA30809.1; -
 DR HSP; P03385; 1MOF.
 DR InterPro; IPR002050; Env polyprotein.
 DR InterPro; IPR008981; F_MuLV bind.
 DR Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 465 COAT PROTEIN GP70.

FT CHAIN 466 668 COAT PROTEIN P15E. (POTENTIAL).
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 668 AA; 74298 MW; 46A606CF2E86CE25 CRC64;
 Query Match 14.2%; Score 408; DB 1; Length 668;
 Best Local Similarity 29.1%; Pred. No. 1.9e-24;
 Matches 120; Conservative 73; Mismatches 158; Indels 62; Gaps 14;
 QY 136 LTRVHGTSPPYKGLDLSKHLTHTRVLSLNTLTGLHEVSAQNPTNCWICLPLNFR 195
 DB 282 ITLVNAGIAPLSTPVPASPRIKGTGNRLNLVQGYTLTNPNKTKDCWLCVSRPP 341
 QY 196 PYVSIPIVPEQWNPSTNTITSVLGLVLSNLTITHTSNTLTCVKFSNTTYTNSQCIRW 255
 DB 342 YVEGIIV--LGNYSNTNPPSPCLSVPHKLTISEVSG-----QGLCIATV 385
 QY 256 TPPTQIVC-----LPSGIFVCGTSAYRC-----INGSSSMCFSLSLVPP 296
 DB 386 PKTHQALCNKTKQKGRGTHYLVAPNGTYWACNTGLTFCISMAVLNWTSD-FCVLTSLNPR 444
 QY 297 MTIYTEQDLYSYISKPRKRVPLPFIYVIGAGVLGALGTGIGGITT-----TOFY 347
 DB 445 IYHEPEYVISHPENKPKFKEDPISLV--ALMLG--GITVGMARNRNRDCGLLETAQF 500
 QY 348 YKLSQELNGMVERVADSLVLODQNSLAAVLQNRALLTAERGTCFLFGECCYY 407
 DB 501 RQLQWAMHTDQALEESISALEKSLTSLSEVQLNRRGLDLFLQEGGLTALKERCCFY 560
 QY 408 VNQSGIVTEKVEIRDIQRAEELNRTGPGWLLSOMW---PW-----ILPFLGLAAAIL 460
 DB 561 ADHTGLVRDNMAKRELKQR-QQLFDSQQDG-LEGFNKSPWFTTILISSIMFLMLLL 618
 QY 461 LLLFGPCIFNLLNVFSSRIEAVKQWEPKQSKTKIYR-RPLDRPASPSDV 512
 DB 619 ILLFGPCILNRLVQVFKDRISVQTLVLTQYQRLQWRLRP---TVSPQLNV 668
 RESULT 19
 ENV_MCFE STANDARD; PRT; 636 AA.
 AC P15073;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
 DE GP70; Spike protein P15E; R protein].
 GN ENV.
 OS Mink cell focus-forming murine leukemia virus.
 OC Viruses; Retroid viruses; Retroviridae.
 OX NCBI_TaxID=11935;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89085614; PubMed=2535909;
 RA Chattopadhyay S.K., Baroud B.M., Holmes K.L., Fredrickson T.N.,
 RA Lander M.R., Morse H.C. III, Hartley J.W.;
 RT "Biologic and molecular genetic characteristics of a unique MCF virus
 RT that is highly leukemogenic in ecotropic virus-negative mice";
 RL Virology 168:90-100 (1989).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M23029; -- NOT_ANNOTATED_CDS.
 DR HSPF; P03385; 1MOF.
 DR InterPro; IPR002050; Env_polypeptide.
 DR InterPro; IPR008981; F_Mulv_bind.
 DR Pfam; PF00429; Env_polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 440 KNOB PROTEIN GP70.
 FT CHAIN 441 620 SPIKE PROTEIN P15E.
 FT CHAIN 621 636 R PROTEIN.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 636 AA; 69113 MW; DF6208F7EA968E2A CRC64;
 Query Match 14.2%; Score 406.5; DB 1; Length 636;
 Best Local Similarity 29.6%; Pred. No. 2.3e-24;
 Matches 104; Conservative 64; Mismatches 144; Indels 39; Gaps 9;
 QY 160 THTRVLSLNTLTGLHEVSAQNPTNCWICLPLNFRPVYIPVPEQWNPSTNTITSVL 219
 DB 281 TGDRLNLVDGAYQALNLTSPDKTQECWLCVAGPYEGVAVLGTYSNHTSAPANCSVA 340
 QY 220 VGPLVSNLEIT-----HTSNLTCVKFSNTTYTNSQCIRWTPPTQIVCLPSGIF 269
 DB 341 SQHKLTLSEVTGQGLCVGAVPKTHQALC---NTTKTSD-----GSYLAAPAGTI 388
 QY 270 FVCGTSAYRCNLGS-----SESMCFSLVFPWMTIYTEQDLYSYISKPRKRVPI---LP 322
 DB 389 WACNTGLTPCLSTVNLNTDYCVLVELWPKVTHSPDYVTIQEPGARFRREPVSLLTA 448
 QY 323 FVIGAGVLGALGTGIGGITT---STQFYKLSQELNGMVERVADSLVLTQDQNSLAAY 379
 DB 449 LLLGLTWGGAAGVGTCTALVATQFQQLQAAVHNDLKEVKSITNLSKSLTSLSEVA 508
 QY 380 LQNRALLDLTAERGTCFLFGECCYYVNSGIVTEKVEKIRDIQRAEELNRTGPW- 438
 DB 509 LQNRRLGLDLFLKEGGLCAALKECCFYADHTGLVRDSMAKLRERLNQRQKLFESGQGW 568
 QY 439 -GLLSQWMPW-----ILPFLGLAAAILLLFGPCIFNLLNVFSSRIEAVK 484
 DB 569 EGLFNR-SPWFTTILSTINGPLVILLLLFGPCILNRLVQVFKDRISVVQ 618
 RESULT 20
 ENV_MLVAV STANDARD; PRT; 669 AA.
 AC P03366;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
 DE GP70; Spike protein P15E; R protein].
 GN ENV.
 OS AKV murine leukemia virus.
 OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11791;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84115072; PubMed=6319746;
 RA Herr W.;
 RT "Nucleotide sequence of AKV murine leukemia virus";
 RL J. Virol. 49:471-478 (1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83090450; PubMed=6294621;

[illegible]

```

QY 373 NSLAIVLQNRALDLTAERGCTCLFLGECVYVNSQSGIVTEKVKIEIRDIORRAEL 432
DB 503 TSLSEVLQNRGLDLFLKEGGLCAALKKECCCFYADHTGLVRDSMAKLERLSQRKLF 562
QY 433 RNTGPM--GLLSQWMPW----ILFPLGLAAILILLLLFGPCIFNLLVNFVSSRIEAVK 484
DB 563 ESQGWFEGLFNK-SPWFTTLLSTIMGLPILILLILLFGPWILNRLVQFVKDRISVVQ 619

RESULT 22
ENV_MLVRD
ID ENV_MLVRD STANDARD; PRT; 665 AA.
AC P14268;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
DE GP76; Spike protein P15E; R protein].
GN ENV
OS Radiation murine leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11787;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87207680; PubMed=3033897;
RA Merregate J., Janowski M., Reddy E.P.;
RT "Nucleotide sequence of a radiation leukemia virus genome.";
RL Virology 158:98-102(1987).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K03363; AAA46519.1; ALT_INIT.
DR PIR; C26183; VCMVVR.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env polypeptide.
DR InterPro; IPR008981; F_MLV bind.
DR Pfam; PF00429; Env polypeptide; 1.
KW Glycoprotein; Coat protein; Polypeptide; Transmembrane; Signal.
FT SIGNAL 1 31
FT CHAIN 32 467
FT CHAIN 468 644
FT CHAIN 645 665
FT TRANSMEM 605 621
FT CARBOHYD 43 43
FT CARBOHYD 199 199
FT CARBOHYD 211 211
FT CARBOHYD 324 324
FT CARBOHYD 356 356
FT CARBOHYD 363 363
FT CARBOHYD 396 396
FT CARBOHYD 400 400
FT CARBOHYD 432 432
FT SEQUENCE 665 AA; 73083 MW; 93B2BFECC9E45984 CRC64;
Query Match
Best Local Similarity 14.0%; Score 401; DB 1; Length 665;
Matches 108; Conservative 66; Mismatches 154; Indels 64; Gaps 9;
QY 163 RLVSLENTTLGLHEVSAQNPTNCWICLIPRPFVSPVPEQNNSTENTTSVLVGP 222
DB 311 RLNLVKGAYQALNLTSPDRQECWCLVSGPPYEGVAVLGTYSNHTSAPANGSVALQH 370
QY 223 LVSNLEIT-----HTSNLTCKFESNTTNTNSQIRWVTPPTQIVCLPSPGIFVC 272
DB 371 KLTLSVITGQGLCVCAVPTQALC-----NTQNTSG-----GSYLAAPAGTWAC 418

```

```

QY 273 GTSAYRCLNGS-----SESCFLSFLVPPMTYITEODLYSYVISKPRKRVPI---LPFVI 325
DB 419 NTGLTPCLSTTVLNTTTCVLEWLPVTHSPSYVTHQFERRGKYKRPVSLTLALL 478
QY 326 GAGVLGALGTGIGGTTTSTQFYKLSQBLNGDMERVADSLVTLQDLNSLAIVLQNRRA 385
DB 479 GGLTWGGAAGIGTGTALVATQQLQAAVHDDLKEVKSITNLEKSLTSLSEVLQNRGG 538
QY 386 LDLLTAERGCTCLFLGECVYVNSQSGIVTEKVKIEIRDIORRA-----BELRNT 435
DB 539 LDLLFLKGGGLCAALKKECCCFYADHTGLVRDSMAKLERLSQRKLFESQGWFERLFG 598
QY 436 GPWGLLSQWMPWILFPLGLAAILILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKXQSKT 495
DB 599 SPW-----FTTLLSTIMGLPILILLILLFGPWILNRLVQFVKDRISVVQALV---LT 647
QY 496 KLYRPLDRPASPRSDVDNDIKGTPPEISAAQ 527
DB 648 QQY-----HQLKSIDPEEMESRE 665

RESULT 23
ENV_FLVSA
ID ENV_FLVSA STANDARD; PRT; 639 AA.
AC P06752;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
DE GP70; Spike protein P15E].
GN ENV
OS Feline leukemia virus (strain C/Sarna).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=103919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86308240; PubMed=3018287;
RA Riedel N., Hoover E.A., Gasper P.M., Nicolson M.O., Mullins J.I.;
RT "Molecular analysis and pathogenesis of the feline aplastic anemia
RL J. Virol. 60:242-250(1986).
CC -/- P-TM: Specific enzymatic cleavages in vivo yield mature proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14331; AAA43049.1; --
DR PIR; A29013; VCMWSA.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env polypeptide.
DR InterPro; IPR008981; F_MLV bind.
DR Pfam; PF00429; Env polypeptide; 1.
KW Coat protein; Glycoprotein; Polypeptide; Signal.
FT SIGNAL 1 33
FT CHAIN 34 442
FT CHAIN 443 639
FT CARBOHYD 35 35
FT CARBOHYD 43 43
FT CARBOHYD 58 58
FT CARBOHYD 299 299
FT CARBOHYD 304 304
FT CARBOHYD 328 328
FT CARBOHYD 331 331
FT CARBOHYD 387 387
FT CARBOHYD 407 407
FT SEQUENCE 639 AA; 71161 MW; A309B64AC8EC74E4 CRC64;
Query Match
Best Local Similarity 13.9%; Score 399.5; DB 1; Length 639;
Matches 108; Conservative 66; Mismatches 154; Indels 64; Gaps 9;
QY 163 RLVSLENTTLGLHEVSAQNPTNCWICLIPRPFVSPVPEQNNSTENTTSVLVGP 222
DB 311 RLNLVKGAYQALNLTSPDRQECWCLVSGPPYEGVAVLGTYSNHTSAPANGSVALQH 370
QY 223 LVSNLEIT-----HTSNLTCKFESNTTNTNSQIRWVTPPTQIVCLPSPGIFVC 272
DB 371 KLTLSVITGQGLCVCAVPTQALC-----NTQNTSG-----GSYLAAPAGTWAC 418

```

HSP; P03390; IAOI.
MGI: 95598; Rv4.
InterPro: IPR002050; Env polyprotein.
InterPro: IPR008981; F Mulv bind.
Pfam: PF00429; Env polyprotein; 1.
Coat protein; Polyprotein.
DOMAIN 288 291
FT CONFLICT 299 299 D -> T (IN REF. 2).
FT CONFLICT 301 302 LT -> SL (IN REF. 2).
FT CONFLICT 305 305 T -> R (IN REF. 2).
FT CONFLICT 307 307 D -> P (IN REF. 2).
FT CONFLICT 309 320 LPETKPPQAGT -> HSHQTSTGGDR (IN REF. 2).
FT CONFLICT 595 595 R -> K (IN REF. 2).
SEQUENCE 679 AA; 74453 MW; 5242BD0D86554039 CRC64;
Query Watch 13.9%; Score 399.5; DB 1; Length 679;
Best Local Similarity 27.1%; Pred. No. 9.1e-24;
Matches 108; Conservative 69; Mismatches 154; Indels 67; Gaps 11;
QY 160 THTRLVSLFNTLTGLHEVSAQNPTNCWICLPINFRPYVSIYPVEQWNNPSTINTSVL 219
DB 319 TGDRLNLVQGVAYLALNMNTPKTCQWCLVSEPPYEGVAVLGDYTKHET----- 370
QY 220 VQPLVSNLEITHSTNLTCKFNSNTTTSQCIRWTTPPTQIVC-----LP 265
DB 371 -AP-----TNCSSRAQHKLTSEVTGGCKLGAVPKTHQALCNTEPTVSSNYLAP 422
QY 266 SGIFFVCGTSAYRCLNGS-----SESCFLSFLVPPMTIYTEQDLYSYISKPRKRVPI- 320
DB 423 EGTLWACSTGLPCLSTTVLNTTDCVLVELWPKYTHPSEVYVTFEPGVAFRRPEVS 482
QY 321 -LPFVIGAGVLGATGTGGITT-----STQFYKLSQELNGDMERVADSLVTLQDQNSL 375
DB 483 LTLALLGLTGWGIAARVGTGTALVATQQFOLOAAMHNDLKAVEESITNLSRTSL 542
QY 376 AAVLQNRALDILLTAERGTCLFLGEECCYYVNSQGVITEKVKIEIRDRIQRAAEELRNT 435
DB 543 SEVVLQNRGLDLLFLKEGGLCAALKECCFYADHTGLVRDSMAKLERLNQOKLPESG 602
QY 436 GPW--GLLSQWMPW----ILPFGIPLAAIILLFPFCIFNLLVNFVSSRIEAVKLOMEP 489
DB 603 QGWFEGLFNR-SPWFTTLISTMGPLIVLLILLFPICILNRLVQVFKDRISVVQALI-- 659
QY 490 KMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQ 527
DB 660 ----LTQQY-----HQLKSIDPEVESRE 679
RESULT 25
ENV_MLVK
ID AC ENV_MLVK STANDARD; PRT; 665 AA.
IC P31794;
DT DT 01-JUL-1993 (Rel. 26, Created)
DT DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE DE GP76; Coat protein P15E; R protein].
GN ENV.
OS Radiation murine leukemia virus (strain Kaplan).
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_Taxid=31689;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333703; PubMed=1629969;
RA Poliquin L., Bergeron D., Fortier J.L., Paquette Y., Bergeron R.,
RA Rassat E.;
RT "Determinants of thymotropism in Kaplan radiation leukemia virus and
RT nucleotide sequence of its envelope region.";
RL J. Virol. 66:5141-5146 (1992).
CC J-1- PPM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```
DR ENBL; M93052; AAA46526.1; -.
DR FIR; B42743; VCMVKA.
DR HSSP; P03385; IMOF.
DR InterPro; IPR002050; Env polyprotein.
DR InterPro; IPR008981; F MuLV bind.
DR Pfam; PF00429; Env polyprotein; 1.
KW - Glycoprotein; Coat protein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 31
FT CHAIN 32 467 KNOB PROTEIN GP76.
FT CHAIN 468 644 SPIKE PROTEIN P15E.
FT CHAIN 645 665 R PROTEIN.
FT TRANSMEM 605 621 POTENTIAL.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 665 AA; 73085 MW; FAISAB6BUC63FOAA CRC64;
Query Match 13.9%; Score 399; DB 1; Length 665;
Best Local Similarity 27.3%; Pred. No. 9.7e-24;
Matches 107; Conservative 66; Mismatches 155; Indels 64; Gaps 9;
QY 163 RLVSLENTLTGLHEVSAQNPTNCICLPINRPVYSIPVPEQNNFSTEINTTSVLVGP 222
DB 311 RLNLVKGAQALNLTSPDRTQECWLCVSGPPYEGVAVLGTYSNHTSAPANCVSASOH 370
QY 223 LVSNLEIT-----HTSNLTCVKFSNTTYTNSQCIRWVTPPTQIVCLPSGIFVYC 272
DB 371 KLTLSVETGRGLCVGAVPKTHQALC---NTQNTSG-----GSYYLAAPAGTIWAC 418
QY 273 GTSAYRCLNGS-----SESMCFSLFVPPMTIYTEODLYSYVISKPNKEVPI---LPFVI 325
DB 419 NTGLTCLSTTVLNLATDYCVLVELWPRVTVHSPSYVYHQFEGRAKYKREPVSLLTALL 478
QY 326 GAGVLGALGTGIGITTSQFYVKLSQELNGDMERVADSLVTIQDOLNSLAADVQLNRRRA 385
DB 479 GGLTMGGIAGVGTGTTALVATQOLAQAVHDDLKEVEKSIITNLEKSLTSLSEVVLQNR 538
QY 386 LDLLTAERGTCFLGRECCYVYVNSQGIYVTEKVKETDRIDRIQRA-----BELRNT 435
DB 539 LDLLFLKEGGLCAALKEECFCYADHTGTVRDSMAKLERLNQKLFESGQGWFERLFPNG 598
QY 436 GPWGLLSQWMPWILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKT 495
DB 599 SPW-----FTTLISTMGFLVILLILLGLPCILNELVQFVKDRISVVQALV-----LT 647
QY 496 KIVRRPLDRPASRSDVDIKGTPPBEISAAQ 527
DB 648 QQY-----HQLKSIDPEMESRE 665
```

Search completed: April 19, 2004, 10:37:52
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:33:58 ; Search time 22 Seconds
(without alignments)
1262.490 Million cell updates/sec

Title: US-10-069-883-1

Perfect score: 2871

Sequence: 1 MALPWHIFLFTVLPSTLT.....PPERISAAQPLLRNSAGSS 538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 2868 | 99.9 | 538 | 4 | US-09-175-928-4 |
| 2 | 2314 | 80.6 | 493 | 4 | US-08-979-847B-106 |
| 3 | 672 | 23.4 | 162 | 4 | US-08-979-847B-109 |
| 4 | 593 | 20.7 | 567 | 1 | US-08-007-282B-2 |
| 5 | 484 | 16.9 | 433 | 3 | US-08-691-563C-87 |
| 6 | 484 | 16.9 | 433 | 4 | US-09-374-768-87 |
| 7 | 484 | 16.9 | 433 | 4 | US-08-979-847B-81 |
| 8 | 420 | 14.6 | 673 | 3 | US-09-075-272-5 |
| 9 | 418.5 | 14.6 | 632 | 4 | US-09-315-127-2 |
| 10 | 418.5 | 14.6 | 632 | 4 | US-09-315-127-3 |
| 11 | 418.5 | 14.6 | 665 | 4 | US-09-309-572-14 |
| 12 | 418.5 | 14.6 | 665 | 4 | US-09-718-096-14 |
| 13 | 418.5 | 14.6 | 1312 | 4 | US-09-554-572-26 |
| 14 | 412 | 14.4 | 182 | 3 | US-08-486-098-107 |
| 15 | 412 | 14.4 | 192 | 3 | US-08-360-107A-117 |
| 16 | 412 | 14.4 | 192 | 3 | US-08-484-223B-107 |
| 17 | 412 | 14.4 | 192 | 3 | US-08-919-597-107 |
| 18 | 412 | 14.4 | 192 | 3 | US-08-475-668A-107 |
| 19 | 412 | 14.4 | 192 | 3 | US-08-485-551A-107 |
| 20 | 412 | 14.4 | 192 | 3 | US-08-471-913A-107 |
| 21 | 412 | 14.4 | 192 | 3 | US-08-485-264A-107 |
| 22 | 412 | 14.4 | 192 | 4 | US-08-474-343A-107 |
| 23 | 412 | 14.4 | 192 | 4 | US-08-470-896-107 |
| 24 | 412 | 14.4 | 192 | 4 | US-08-485-546A-107 |
| 25 | 411.5 | 14.3 | 638 | 3 | US-09-376-781-6 |
| 26 | 405.5 | 14.1 | 654 | 4 | US-09-315-127-11 |
| 27 | 405.5 | 14.1 | 654 | 4 | US-09-315-127-12 |
| 28 | 403.5 | 14.1 | 660 | 3 | US-09-111-085-2 |
| 29 | 403.5 | 14.1 | 660 | 3 | US-09-376-781-5 |
| 30 | 400.5 | 13.9 | 645 | 4 | US-09-315-127-8 |
| 31 | 400.5 | 13.9 | 645 | 4 | US-09-315-127-9 |
| 32 | 390.5 | 13.6 | 657 | 3 | US-09-111-085-4 |
| 33 | 390.5 | 13.6 | 657 | 3 | US-09-376-781-4 |
| 34 | 389.5 | 13.6 | 667 | 4 | US-09-315-127-5 |
| 35 | 389.5 | 13.6 | 667 | 4 | US-09-315-127-6 |
| 36 | 389 | 13.5 | 232 | 4 | US-09-555-352-10 |
| 37 | 386.5 | 13.5 | 656 | 3 | US-09-376-781-3 |
| 38 | 370 | 12.9 | 67 | 1 | US-08-586-878A-49 |
| 39 | 370 | 12.9 | 67 | 1 | US-08-721-489-3 |
| 40 | 323 | 11.3 | 196 | 2 | US-08-484-126-7 |
| 41 | 323 | 11.3 | 196 | 2 | US-09-374-909-7 |
| 42 | 291 | 10.1 | 152 | 4 | US-09-120-653D-8 |
| 43 | 287 | 10.0 | 77 | 4 | US-08-979-847B-103 |
| 44 | 285.5 | 9.9 | 227 | 4 | US-09-904-615-126 |
| 45 | 213.5 | 7.4 | 486 | 1 | US-07-672-483-2 |
| 46 | 210 | 7.3 | 486 | 1 | US-08-259-451-13 |
| 47 | 187.5 | 6.5 | 488 | 1 | US-07-672-483-1 |
| 48 | 186 | 6.5 | 257 | 4 | US-08-778-717-13 |
| 49 | 180 | 6.3 | 173 | 1 | US-07-743-518-26 |
| 50 | 177.5 | 6.2 | 255 | 4 | US-08-778-717-11 |
| 51 | 174.5 | 6.1 | 178 | 2 | US-08-259-451-15 |
| 52 | 170 | 5.9 | 176 | 2 | US-08-484-126-8 |
| 53 | 170 | 5.9 | 176 | 2 | US-09-374-909-8 |
| 54 | 155 | 5.4 | 160 | 1 | US-08-014-153D-8 |
| 55 | 145.5 | 5.1 | 681 | 3 | US-08-760-615-6 |
| 56 | 143 | 5.0 | 681 | 3 | US-09-336-910A-2 |
| 57 | 141 | 4.9 | 681 | 3 | US-08-760-615-4 |
| 58 | 140 | 4.9 | 49 | 2 | US-08-117-952-792 |
| 59 | 139.5 | 4.9 | 568 | 3 | US-09-160-065-2 |
| 60 | 137.5 | 4.8 | 142 | 1 | US-08-014-153D-6 |
| 61 | 125.5 | 4.4 | 86 | 1 | US-08-014-153D-21 |
| 62 | 121 | 4.2 | 596 | 4 | US-09-345-473B-21 |
| 63 | 120.5 | 4.2 | 99 | 1 | US-08-014-153D-46 |
| 64 | 116.5 | 4.1 | 75 | 3 | US-09-042-012-8 |
| 65 | 116.5 | 4.1 | 75 | 3 | US-09-305-086-2 |
| 66 | 116.5 | 4.1 | 75 | 3 | US-09-457-324-8 |
| 67 | 111 | 3.9 | 49 | 2 | US-08-117-952-791 |
| 68 | 109 | 3.8 | 27 | 4 | US-09-904-615-128 |
| 69 | 106 | 3.7 | 63 | 1 | US-08-014-153D-48 |
| 70 | 105.5 | 3.7 | 676 | 3 | US-08-760-615-2 |
| 71 | 105 | 3.7 | 63 | 1 | US-08-014-153D-49 |
| 72 | 104.5 | 3.6 | 676 | 3 | US-09-650-086A-2 |
| 73 | 104 | 3.6 | 676 | 3 | US-08-447-925-6 |
| 74 | 104 | 3.6 | 602 | 3 | US-09-357-490-15 |
| 75 | 104 | 3.6 | 854 | 4 | US-09-206-551-16 |
| 76 | 104 | 3.6 | 855 | 2 | US-08-482-090-12 |
| 77 | 104 | 3.6 | 856 | 2 | US-08-481-700B-8 |
| 78 | 104 | 3.6 | 856 | 2 | US-09-007-383-16 |
| 79 | 103 | 3.6 | 852 | 4 | US-08-811-682-3 |
| 80 | 99.5 | 3.5 | 862 | 4 | US-09-206-551-15 |
| 81 | 99 | 3.4 | 489 | 2 | US-08-484-126-1 |
| 82 | 99 | 3.4 | 489 | 2 | US-09-374-909-1 |
| 83 | 99 | 3.4 | 863 | 1 | US-08-325-547-4 |
| 84 | 98.5 | 3.4 | 1495 | 4 | US-08-522-726B-1 |
| 85 | 98.5 | 3.4 | 1495 | 4 | US-09-337-384-1 |
| 86 | 98 | 3.4 | 453 | 2 | US-08-484-126-3 |
| 87 | 98 | 3.4 | 453 | 2 | US-09-374-909-3 |
| 88 | 96 | 3.3 | 1239 | 1 | US-08-026-138B-3 |
| 89 | 95.5 | 3.3 | 832 | 4 | US-09-206-551-19 |
| 90 | 95 | 3.3 | 63 | 1 | US-08-014-153D-50 |
| 91 | 95 | 3.3 | 850 | 4 | US-09-206-551-10 |
| 92 | 95 | 3.3 | 850 | 4 | US-09-206-551-52 |
| 93 | 95 | 3.3 | 875 | 4 | US-09-206-551-14 |
| 94 | 95 | 3.3 | 1148 | 4 | US-09-106-075A-90 |
| 95 | 94 | 3.3 | 63 | 1 | US-08-014-153D-47 |
| 96 | 94 | 3.3 | 546 | 4 | US-09-907-794A-250 |
| 97 | 94 | 3.3 | 546 | 4 | US-09-905-125A-250 |
| 98 | 94 | 3.3 | 546 | 4 | US-09-902-775A-250 |
| 99 | 93 | 3.2 | 1422 | 4 | US-08-469-260A-83 |
| 100 | 93 | 3.2 | 1422 | 4 | US-08-488-446-83 |

ALIGNMENTS

RESULT 1
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R. A.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Metberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B-AJ172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-175-928-4

Query Match 99.9%; Score 2868; DB 4; Length 538;
Best Local Similarity 99.8%; Pred. No. 1.2e-285;
Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSSPYQEFLLWRMQRPGNIDAPSYRSLKGTG 60
DB 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSSPYQEFLLWRMQRPGNIDAPSYRSLKGTG 60

QY 61 TFTATHMPCNCHSATLCMHANTHYTGKMINPCSCGGLGVTVCTWYFTQTGMSDGGV 120
DB 61 TFTATHMPCNCHSATLCMHANTHYTGKMINPCSCGGLGVTVCTWYFTQTGMSDGGV 120

QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180

QY 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTTSVLGVLNLEIHTHTNLTCVKF 240
DB 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTTSVLGVLNLEIHTHTNLTCVKF 240

QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSCMFLSLVPPMTIY 300
DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSCMFLSLVPPMTIY 300

QY 301 TEQDLYSVISKPRNKRPVLPVIGAGVLGALGTGGITSTQFYKLSQELNGDMER 360
DB 301 TEQDLYSVISKPRNKRPVLPVIGAGVLGALGTGGITSTQFYKLSQELNGDMER 360

QY 361 VADSVLTQDLNLSLAAVLNRRALDLTAERGGTCLFLGEECCYYVNGSIGVTEKVE 420
DB 361 VADSVLTQDLNLSLAAVLNRRALDLTAERGGTCLFLGEECCYYVNGSIGVTEKVE 420

QY 421 IRDRIQRAEELRTGPGWLLSQWMPILFLGPLAAIILLILFGPCIFNLVNVSSRI 480
DB 421 IRDRIQRAEELRTGPGWLLSQWMPILFLGPLAAIILLILFGPCIFNLVNVSSRI 480

QY 481 EAVKLQWEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS 538
DB 481 EAVKLQWEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS 538

RESULT 2

US-08-979-847B-106
; Sequence 106, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; KAMNOS-BACCALA, GLAUCIA
; KURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/979,847B
; FILING DATE: 26-NO. 6582703-1997
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 493 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-08-979-847B-106

Query Match 80.6%; Score 2314; DB 4; Length 493;
Best Local Similarity 88.4%; Pred. No. 9.3e-229;
Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSSPYQEFLLWRMQRPGNIDAPSYRSLKGTG 60
DB 1 MALPYHIFLFTVLLPSTLTAPPCCRCMTSSSPYQEFLLWRMQRPGNIDAPSYRSLKGTG 60

QY 61 TFTATHMPCNCHSATLCMHANTHYTGKMINPCSCGGLGVTVCTWYFTQTGMSDGGV 120
DB 61 TFTATHMPCNCHSATLCMHANTHYTGKMINPCSCGGLGVTVCTWYFTQTGMSDGGV 120

QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180

QY 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTTSVLGVLNLEIHTHTNLTCVKF 240
DB 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTTSVLGVLNLEIHTHTNLTCVKF 240

QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSCMFLSLVPPMTIY 300
DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSCMFLSLVPPMTIY 300

QY 301 TEQDLYSVISKPRKRVPIIPFVIGAGVGLGALGTGIGGITTSTQFYVYKLSQELNGDWER 360
Db 301 TEQDLYNHVVPKPNKRVPIIPFVIRAGVGLGELGTGIGSITTSTQFYVYKLSQELNGDWEQ 360
QY 361 VADSLVTLODQNSLAADVVLQNRALDILITBARGTCLFLGEECCYVYVNSGSIIVTEKYKE 420
Db 361 VTDSLVTLODQNSLAADVVLQNRALDILITAKRGTCCLFLGEECCYVYVNSGSIIVTEKYKE 420
QY 421 IRRDIOREAEELRNTPGMLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
Db 421 IRRDIOREAEELRNTPGMLLSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQ 492
Db 481 EAVKLQWLQME 492

RESULT 3

US-08-979-847B-109
; Sequence 109, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARAMHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-Nov. 6582703-1997
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 109:

Query Match 23.4%; Score 672; DB 4; Length 162;
Best Local Similarity 82.1%; Pred. No. 8.6e-61;
Matches 133; Conservative 6; Mismatches 19; Indels 4; Gaps 1;

QY 381 QNRALDLITARGGTCTFLGEECCYVYVNSGSIIVTEKYKEIRDRIORRAEELNTPGWL 440
Db 1 QNRALDLITAKRGGTCTFLGEECCYVYVNSGSIIVTEKYKEIXDRIXCRAEDLQNTAPWL 60
QY 441 LSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRIEAVK----LQMEPKMQSKTK 496
Db 61 LSQWMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRIEAVKQLQIVLQMEPKMQSKTK 120
QY 497 IYRPLDRPASPRSDYNDIKGTPPEISAAQPLLRPNSSAGSS 538
Db 121 IYRPLDRPARLCSVDYNDIEVTPPEEISTAQPLLLHNSVSS 162

RESULT 4

US-08-007-282B-2
; Sequence 2, Application US/08007282B
; Patent No. 5403582
; GENERAL INFORMATION:

; APPLICANT: NAZERIAN, KEYVAN
; APPLICANT: CALVERT, JAY G.
; APPLICANT: WITTER, RICHARD L.
; APPLICANT: YANAGIDA, NOBORU
; TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
; TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
; TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 810 Gatehouse Road Suite 500 East
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,282B
FILING DATE: 19930121
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR., GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1644-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 567 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-007-282B-2

Query Match 20.7%; Score 593.5; DB 1; Length 567;

Best Local Similarity 33.5%; Pred. No. 8.1e-52;
Matches 154; Conservative 75; Mismatches 176; Indels 55; Gaps 13;
QY 66 THMPRNCYHSATLCMHANTHYWT-----GMINPSPGGLGVTCVCTYFT 110
Db 121 TQWSTCYEXTQECTLLGKTYFTAILQKTKLGSYEDGPNKLIQASCTGTGKPCVCDPVA 180
QY 111 QTGMSDGGGVQDQAREKHVKVVISOLTRVHGTSSPYKGLDLSKLHETLRTHRLVSLFNT 170
Db 181 PVVYSDGGGFTDIREESVRERLEEIRHSYPSVQTHPLALPR-----SRGVLDPO 232

QY 171 T---LTGLHEV- SAQNP---TNCWICLPLNFRPYVSIIPVPEQNNFSTINTTSLVGLPL 223
Db 233 TSDILEATHOVLNATNPKLAENCWLCMTLTGTIPAAIP-----TNGNVTLDGNCSLSPF 287
QY 224 VSN---LBIT---HTSNLTCVKSNTYTTNSQCIRWVTPQTQ-----IVCLPSSGIF 269
Db 288 GGNPPGSDVSCVAGRADNRTGIPGVYHF-TNCTSIQEVNETSQMGNLRTLCPPPGHV 346
QY 270 FVCGTS-AYRCLNGSSESCFLSFLVPPMTIYTEQD---LYSYVISKPRNR-VFPLPFV 324
Db 347 FVCGNNMAYTLPKNWIGLCILASIVPDISIISGEEPIPSIEVTARRHKRAVQFILL 406
QY 325 IGAVLUGALGTGGTITSTQFYKLSQELNGDMERVADSLVTLQDOLNSLAADVQLNRR 384
Db 407 VGLIGSGATLAGTGLGVSVHTKLSNQLIISDVQALSGTINDLQDQDLSLAEVQLNRR 466
QY 385 ALDLTAERGGTCLFLGEECCYVYVQSGIVTEKVEIRDRIQRRABEELNTPGWLGLSOW 444
Db 467 GUDLLTAEGGICLALQEKCCFVANKSGIVRKIKLQEDLIERKALYDNPWLSGLNGF 526
QY 445 MPWLPFLGPLAAIILLLLFGPCIFNLVNFVSSRIEAVK 484
Db 527 LPYLLPLGLGLLFLTGPCKIMKTLTRIHDKIQAVK 566

RESULT 5

US-08-691-563C-87
; Sequence 87, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-691-563C-87

Query Match 16.9%; Score 484; DB 3; Length 433;
Best Local Similarity 53.0%; Pred. No. 9.3e-41;

Matches 97; Conservative 27; Mismatches 41; Indels 18; Gaps 4;
QY 332 ALGTGIGGTTSTQFYKLSQELNGDMERVADSLVTLQDOLNSLAADVQLNRRALDLITA 391
Db 4 ATGTGIAGLSTLSYHTLSKNFSDLSQELIMKSLTQSDLSLAATLQNRGPHLLTA 63
QY 392 ERGTGCLFLGEECCYVYVQSGIVTEKVEIRDRIQRRABEELN-----TGPWGLLSQWM 445
Db 64 EKGLCTFLGEECCFYNQSGI-----VRDATWHLQERASDIRQCLNSYTNLW---SWA 115
QY 446 PWLPFLGPLAAIILLLLFGPCIFNLVNFVSSRIEAVK-----LQEPKMQSKTKYRRP 501
Db 116 TWLPLFGPWAAILLLLTGFCIFKLLVFKVSSRIEAIKQLMVLOMEPQMSSTNFFYQGP 175
QY 502 LDR 504
Db 176 LER 178
RESULT 6
US-09-374-766-87
; Sequence 87, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/374,766
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-374-766-87

Query Match 16.9%; Score 484; DB 4; Length 433;
Best Local Similarity 53.0%; Pred. No. 9.3e-41;
Matches 97; Conservative 27; Mismatches 41; Indels 18; Gaps 4;

QY 332 ALGTGIGGTTSTQFYKLSQELNGDMERVADSLVTLQDOLNSLAADVQLNRRALDLITA 391

Db 386 AVSNGCLGQVQDQKHLNCTQNTQPNKGGQYLVPPIDTV-----WACNTGLTPCI 438
QY 281 N-----CSSSMCFELSLVPPMTIYTQDLYSYVSKPRNKRVP1---LPFVIGAGVLGA- 332
Db 439 SMSVFNSSKDFCLVQLPRLLYHDDSSFLDKPEHRVRWRKREPITLTLAVLGLGJGAAG 498
QY 333 LGTIGIGITTTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAADVQLQNRALDLLTAE 392
Db 499 VGTGTAALIQTPRYFEELTAMDLDLRAEHSITKLESLTSLSEVLQNRRLDILLFLK 558
QY 393 RGTCLFLGEECCYYNQSGIVTEKVEIRDRIOREAEELRNTGPNGLLSQMW---PWIL 449
Db 559 EGLCAALKEECFCYVDHSGVVDKSWAKURERLDIFQREESKQGM---FESFNKSPMLT 616
QY 450 PFL-----GPIAAILLLFGPCIFNLLVNFVSSRIEAVKLQ 487
Db 617 TLLSTIAGPLIILLLLLTGFCILNKLVAFIRERINAVQMV 658

RESULT 9
US-09-315-127-2
; Sequence 2, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315,127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-315-127-2

Query Match 14.6%; Score 418.5; DB 4; Length 632;
Best Local Similarity 30.2%; Pred. No. 9.5e-34;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THRLVSLFNTTLTGLHEVSAQNPNTNCWICLPLNFRPYVSIPIVPEQNNFSTEINTSVL 219
Db 277 TENRLNLVDGAYQALNLTSPDKTQECWLCIVAGPPYEGVAVLGTYSNHTSAPANCVA 336
QY 220 VGPLVSNLEIT-----HTSNLTCVKFSNTTYTNSQCIRKWTPPTQIVCLPSGIF 269
Db 337 SQHKLTLSEVTGQGLCIGAVPKTHQALC-----NTQTSSR-----GSYLVAPTGM 384
QY 270 FVCGTSAYRCLNGS-----SESMLFSLVPPMTIYTQDLYSYVSKPRNKRVP1---LP 322
Db 385 WACSTGLTPCISTILNLTDDYCVLVELMPRVYHSPSYVGLFERSNRHKREPVSLTLA 444
QY 323 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSLAADV 379
Db 445 LLLGGLTWGGIAGTGTALMATQFQQLAAVQDQDLREVEKSIKSLTSLSEV 504
QY 380 LQNRALDLLTAERGCTCLFGECCYYVNSGIVTEKVEIRDRIOREAEELRNTGPN- 438
Db 505 LQNRGLDILLKGGGLCAALKEECFCYVDHGTGLVRDSMAKLERLNQKLFESTQGW 564
QY 439 -GLLSQMPW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 484
Db 565 EGLFNR-SPWFTTLISTMGPLVILLMLLFGPCILNRLVQFVKDRISVVQ 614

RESULT 10
US-09-309-572-14
; Sequence 14, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: env protein
US-09-309-572-14

Query Match 14.6%; Score 418.5; DB 4; Length 665;
Best Local Similarity 30.2%; Pred. No. 1e-33;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315,127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3
; LENGTH: 632
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQ ID NO.2,
; OTHER INFORMATION: envelope protein produced by retroviral vector of
; OTHER INFORMATION: seq. id no.1
US-09-315-127-3

Query Match 14.6%; Score 418.5; DB 4; Length 632;
Best Local Similarity 30.2%; Pred. No. 9.5e-34;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THRLVSLFNTTLTGLHEVSAQNPNTNCWICLPLNFRPYVSIPIVPEQNNFSTEINTSVL 219
Db 277 TENRLNLVDGAYQALNLTSPDKTQECWLCIVAGPPYEGVAVLGTYSNHTSAPANCVA 336
QY 220 VGPLVSNLEIT-----HTSNLTCVKFSNTTYTNSQCIRKWTPPTQIVCLPSGIF 269
Db 337 SQHKLTLSEVTGQGLCIGAVPKTHQALC-----NTQTSSR-----GSYLVAPTGM 384
QY 270 FVCGTSAYRCLNGS-----SESMLFSLVPPMTIYTQDLYSYVSKPRNKRVP1---LP 322
Db 385 WACSTGLTPCISTILNLTDDYCVLVELMPRVYHSPSYVGLFERSNRHKREPVSLTLA 444
QY 323 FVIGAGVLGALGTGIGITT---STQFYKLSQELNGDMERVADSLVTLQDQNSLAADV 379
Db 445 LLLGGLTWGGIAGTGTALMATQFQQLAAVQDQDLREVEKSIKSLTSLSEV 504
QY 380 LQNRALDLLTAERGCTCLFGECCYYVNSGIVTEKVEIRDRIOREAEELRNTGPN- 438
Db 505 LQNRGLDILLKGGGLCAALKEECFCYVDHGTGLVRDSMAKLERLNQKLFESTQGW 564
QY 439 -GLLSQMPW-----ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 484
Db 565 EGLFNR-SPWFTTLISTMGPLVILLMLLFGPCILNRLVQFVKDRISVVQ 614

RESULT 11
US-09-309-572-14
; Sequence 14, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Moloney murine leukemia virus
; FEATURE:
; OTHER INFORMATION: env protein
US-09-309-572-14

Query Match 14.6%; Score 418.5; DB 4; Length 665;
Best Local Similarity 30.2%; Pred. No. 1e-33;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THTRLVSLNTLTGLHEVSAQNPTNCWICLPLNRPYVSIPTVPMQNNFSTEINTSVL 219
DB 310 TENRLNLVDGAYQALNLTSPDKTQECWCLVAGPPYEGVAVLGTYSNHTSAPANCVA 369
QY 220 VGPLVSNLEIT-----HTSNLTCKVFNNTYTNQSCIRWVTPTQIVCLPSGIF 269
DB 370 SQKLTLSVETGQGLCIGAVPKTHQALC-----NTQTSSR-----GSYLVAPTGT 417
QY 270 FVCGTSAYRCLNGS-----SESFCFLSFLVPPMTIYTEQDLYSYVSKPRNKRVP 322
DB 418 WACSTGLTPTCISTTILNLTDDYCVLVELWPRVTHSPSYVGLFERSNRHREPVSLT 477
QY 323 FVIGAGVLGALGTGIGGITT-----STQFYKLSQELNGDMERVADSLVTLQDLSLA 379
DB 478 LLLGLTGMGIAAGTGTATMALMATQFQQLAAVQDLDREVEKSIINLEKSLTSLSEV 537
QY 380 LQNRALDILLTAERGTCFLGEBECYVYVNSQGIIVTEKVEIRDRIOREAEELNTGP 438
DB 538 LQNRGLDILLFKEGGLCAALKECCFYADHTGLVRDSMAKLERLNQKLFESTQGW 597
QY 439 -GLLSQWMPW-----ILPFLGPLAAIILLFPGCFLNLLVNFVSSRIEAVK 484
DB 598 EGLFNR-SPWFTTLISTINGPLIVLLMILLFGPCILNRLVQFVKDRISVVQ 647

RESULT 12

US-09-718-096-14

; Sequence 14, Application US/09718096

; Patent No. 6589763

; GENERAL INFORMATION:

; APPLICANT: Von Laer, Melike-Dorothee

; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV

; FILE REFERENCE: 35-195

; CURRENT APPLICATION NUMBER: US/09/718,096

; PRIOR FILING DATE: 2000-11-22

; PRIOR APPLICATION NUMBER: DE 19856463.5

; PRIOR FILING DATE: 1998-11-26

; PRIOR APPLICATION NUMBER: EP 99250415.9

; PRIOR FILING DATE: 1999-11-25

; PRIOR APPLICATION NUMBER: US 09/309,572

; PRIOR FILING DATE: 1999-05-11

; NUMBER OF SEQ ID NOS: 26

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 14

; LENGTH: 665

; TYPE: PRT

; ORGANISM: Moloney murine leukemia virus

; FEATURE:

; OTHER INFORMATION: env protein

US-09-718-096-14

Query Match 14.6%; Score 418.5; DB 4; Length 665;
Best Local Similarity 30.2%; Pred. No. 1e-33;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THTRLVSLNTLTGLHEVSAQNPTNCWICLPLNRPYVSIPTVPMQNNFSTEINTSVL 219
DB 310 TENRLNLVDGAYQALNLTSPDKTQECWCLVAGPPYEGVAVLGTYSNHTSAPANCVA 369
QY 220 VGPLVSNLEIT-----HTSNLTCKVFNNTYTNQSCIRWVTPTQIVCLPSGIF 269
DB 370 SQKLTLSVETGQGLCIGAVPKTHQALC-----NTQTSSR-----GSYLVAPTGT 417
QY 270 FVCGTSAYRCLNGS-----SESFCFLSFLVPPMTIYTEQDLYSYVSKPRNKRVP 322
DB 418 WACSTGLTPTCISTTILNLTDDYCVLVELWPRVTHSPSYVGLFERSNRHREPVSLT 477
QY 323 FVIGAGVLGALGTGIGGITT-----STQFYKLSQELNGDMERVADSLVTLQDLSLA 379
DB 478 LLLGLTGMGIAAGTGTATMALMATQFQQLAAVQDLDREVEKSIINLEKSLTSLSEV 537
QY 380 LQNRALDILLTAERGTCFLGEBECYVYVNSQGIIVTEKVEIRDRIOREAEELNTGP 438

DB 538 LQNRGLDILLFKEGGLCAALKECCFYADHTGLVRDSMAKLERLNQKLFESTQGW 597
QY 439 -GLLSQWMPW-----ILPFLGPLAAIILLFPGCFLNLLVNFVSSRIEAVK 484
DB 598 EGLFNR-SPWFTTLISTINGPLIVLLMILLFGPCILNRLVQFVKDRISVVQ 647
RESULT 13
US-09-554-572-26
; Sequence 26, Application US/09554572
; Patent No. 6573091
; GENERAL INFORMATION:
; APPLICANT: NATURE TECHNOLOGY, INC.
; TITLE OF INVENTION: CHIMERIC VIRAL PACKAGING SIGNAL WITHOUT GAG GENE
; TITLE OF INVENTION: SEQUENCES
; FILE REFERENCE: 228.00030201
; CURRENT APPLICATION NUMBER: US/09/554,572
; CURRENT FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Murine leukemia virus
US-09-554-572-26

Query Match 14.6%; Score 418.5; DB 4; Length 1312;
Best Local Similarity 30.2%; Pred. No. 3.2e-33;
Matches 106; Conservative 63; Mismatches 143; Indels 39; Gaps 9;

QY 160 THTRLVSLNTLTGLHEVSAQNPTNCWICLPLNRPYVSIPTVPMQNNFSTEINTSVL 219
DB 957 TENRLNLVDGAYQALNLTSPDKTQECWCLVAGPPYEGVAVLGTYSNHTSAPANCVA 1016
QY 220 VGPLVSNLEIT-----HTSNLTCKVFNNTYTNQSCIRWVTPTQIVCLPSGIF 269
DB 1017 SQKLTLSVETGQGLCIGAVPKTHQALC-----NTQTSSR-----GSYLVAPTGT 1064
QY 270 FVCGTSAYRCLNGS-----SESFCFLSFLVPPMTIYTEQDLYSYVSKPRNKRVP 322
DB 1065 WACSTGLTPTCISTTILNLTDDYCVLVELWPRVTHSPSYVGLFERSNRHREPVSLT 1124
QY 323 FVIGAGVLGALGTGIGGITT-----STQFYKLSQELNGDMERVADSLVTLQDLSLA 379
DB 1125 LLLGLTGMGIAAGTGTATMALMATQFQQLAAVQDLDREVEKSIINLEKSLTSLSEV 1184
QY 380 LQNRALDILLTAERGTCFLGEBECYVYVNSQGIIVTEKVEIRDRIOREAEELNTGP 438
DB 1185 LQNRGLDILLFKEGGLCAALKECCFYADHTGLVRDSMAKLERLNQKLFESTQGW 1244
QY 439 -GLLSQWMPW-----ILPFLGPLAAIILLFPGCFLNLLVNFVSSRIEAVK 484
DB 1245 EGLFNR-SPWFTTLISTINGPLIVLLMILLFGPCILNRLVQFVKDRISVVQ 1294

RESULT 14

US-08-486-099-107

; Sequence 107, Application US/08486099

; Patent No. 6013263

; GENERAL INFORMATION:

; APPLICANT: Bolognesi, Dani P.

; APPLICANT: Matthews, Thomas J.

; APPLICANT: Wild, Carl T.

; APPLICANT: Barney, Shawn O.

; APPLICANT: Lambert, Dennis M.

; APPLICANT: Petteway, Stephen R.

; APPLICANT: Langlois, Alphonse J.

; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF

; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS

; TITLE OF INVENTION: B VIRUS TRANSMISSION

; NUMBER OF SEQUENCES: 209

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,099
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-486-099-107

Query Match 14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLCALGTGGITTTQFYVYKLSQELNGDMERVADSLVTLQDOLNSLAA 377
DB 2 IQILPFLVGLGITVAVSTGAGLGVSITQYTKLSHQLISDVQAISSTIQDQDVSIAE 61
QY 378 VVLQNRALDLLTAERGTCFLGEECCYVYVNSGIVTEKVKIQRRAEELRNTGP 437
DB 62 VVLQNRGLDLLTAEOGGICLALQKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQMPWILFLGLPLAAILLLFGPCIFENLLVNFVSSRIEAVK 484
DB 122 WTSFHGLFVYVWPLGLLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 15
US-08-360-107A-117
Sequence 117, Application US/08360107A
Patent No. 6017536
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 149
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/360,107A
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-360-107A-117

Query Match 14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLCALGTGGITTTQFYVYKLSQELNGDMERVADSLVTLQDOLNSLAA 377
DB 2 IQILPFLVGLGITVAVSTGAGLGVSITQYTKLSHQLISDVQAISSTIQDQDVSIAE 61
QY 378 VVLQNRALDLLTAERGTCFLGEECCYVYVNSGIVTEKVKIQRRAEELRNTGP 437
DB 62 VVLQNRGLDLLTAEOGGICLALQKCCFYANKSGIVRDKIKNLQDQLERRRRQLIDNPF 121
QY 438 WGLLSQMPWILFLGLPLAAILLLFGPCIFENLLVNFVSSRIEAVK 484
DB 122 WTSFHGLFVYVWPLGLLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 16
US-08-484-223B-107
Sequence 107, Application US/08484223B
Patent No. 6020459
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 245
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,223B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-484-223B-107

Query Match          14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Fred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;

QY      318  VPILPFVIGAGVLGALGTGIGGITTSTQFYKYSQELNGDMERVADSLVTLQDQNSLAA 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2    IQLPLFVLGIGTITAVSTGAAGLGSVITQYTKLSHQLISDVQAISSTIQDLQDQVDSLAE 61

QY      378  VVLQNRALDLLTAERGTCFLPEBECYCVYNSGIVTEKVKIRDIRQRAEELRNTGP 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62  VVLQNRGLDLLTAEGGIGICLALOEKCCFVANKSGIVRDKIKVLQDDELRRLQIDNPF 121

QY      438  WGLLSQMPFWILPFGLPLAAIILLFLFGPCIFNLLVNFVSSRIEAYK 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122  WTSPFGLPYMPLGLPLCLLAVLSEGPILFNKLTWFIKHQIESIO 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 17
US-08-919-597-107
; Sequence 107, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-107

Query Match          14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0

Qy      318  VFLPFVITGAGVLGALTGIGITTSQFYKYSQELNGDMERVADSLVTLQQLNSLAA 377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2    IQLIFLVGLGITTAVTSGAAGLGSVITQYTKLSHQLISDVQAISSTIQDLQQVDSLAE 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      378  VYLQNRRLDILLTAERGTCFLFGEBCYYVNSGIVTEKVKGIIRDRIQRRAELELNTGP 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62  VYLQNRGLDILLTAERGICLALQEKCFYANKSGIVRDKIKLQDLLERRRLQIDNPF 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      438  WGLLSQWFWPLFLFGLPLAAIILLLLGPCIFNLLNVNFSRIBAVK 494
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122  WTSFGHFLPYVWPLFGLPLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 168
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

US-08-475-668A-107

Query Match 14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;

QY 318 VPIPLFVIGVGLGALGTGIGGTTSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAA 377
DB 2 IQIPLFVGLGTTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQDQVDSLAE 61
QY 378 VVLQNRALDLLTAERGTCFLFGECCYVYVNSGIVTEKVEIRDRIOQRAEELRNTGP 437
DB 62 VVLQNRGLDLLTAEOGGICLALQEKCCFYANKSGIVRDKIKNLQDQLERRRRQLINDPFF 121
QY 438 WGLLSOWMPWILPFLGLAAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGLLCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 19

US-08-485-551A-107
Sequence 107, Application US/08485551A
Patent No. 6068973

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,551A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 107:

SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-485-551A-107

Query Match 14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;

QY 318 VPIPLFVIGVGLGALGTGIGGTTSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAA 377

DB 2 IQIPLFVGLGTTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQDQVDSLAE 61
QY 378 VVLQNRALDLLTAERGTCFLFGECCYVYVNSGIVTEKVEIRDRIOQRAEELRNTGP 437
DB 62 VVLQNRGLDLLTAEOGGICLALQEKCCFYANKSGIVRDKIKNLQDQLERRRRQLINDPFF 121
QY 438 WGLLSOWMPWILPFLGLAAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGLLCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 20

US-08-471-913A-107
Sequence 107, Application US/08471913A
Patent No. 6093794

GENERAL INFORMATION:

APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.

TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 214

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,913A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 107:

SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein

US-08-471-913A-107

Query Match 14.4%; Score 412; DB 3; Length 192;

Best Local Similarity 44.3%; Pred. No. 6e-34; Indels 0; Gaps 0;
Matches 74; Conservative 42; Mismatches 51;

QY 318 VPIPLFVIGVGLGALGTGIGGTTSTQFYKLSQELNGDMERVADSLVTLQDQLNSLAA 377

DB 2 IQIPLFVGLGTTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQDQVDSLAE 61

QY 378 VVLQNRALDLLTAERGTCFLFGECCYVYVNSGIVTEKVEIRDRIOQRAEELRNTGP 437

DB 62 VVLQNRGLDLLTAEOGGICLALQEKCCFYANKSGIVRDKIKNLQDQLERRRRQLINDPFF 121

QY 438 WGLLSQWMPILPFLGALAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 21
US-08-485-264A-107
; Sequence 107, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-485-264A-107

Query Match 14.4%; Score 412; DB 3; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLGALGTGIGITTTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
DB 2 IQILPFLVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDLDQDQVDSLAE 61
QY 378 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
DB 62 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
QY 438 WGLLSQWMPILPFLGALAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 22
US-08-474-349A-107

; Sequence 107, Application US/08474349A
; Patent No. 6333395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/474,349A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-474-349A-107

Query Match 14.4%; Score 412; DB 4; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLGALGTGIGITTTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
DB 2 IQILPFLVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDLDQDQVDSLAE 61
QY 378 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
DB 62 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
QY 438 WGLLSQWMPILPFLGALAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 23
US-08-470-896-107

; Sequence 107, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/470,896
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-107

Query Match 14.4%; Score 412; DB 4; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLGALGTGIGITTTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
DB 2 IQILPFLVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDLDQDQVDSLAE 61
QY 378 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
DB 62 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
QY 438 WGLLSQWMPILPFLGALAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 24
US-08-470-896-107

; Sequence 107, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/470,896
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-107

Query Match 14.4%; Score 412; DB 4; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;
QY 318 VPILPFVIGAGVLGALGTGIGITTTSTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
DB 2 IQILPFLVGLGITTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDLDQDQVDSLAE 61
QY 378 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
DB 62 VVLQNRRLDLLTAERGCTFLGECCEVYVNSGIVTEKVKIEIRDRIRRAEELRNTGP 437
QY 438 WGLLSQWMPILPFLGALAIILLLFGPCIFNLVNFVSSRIEAVK 484
DB 122 WTSFHGFLPYVMPFLGCLLLVLSFGPIIFNKMTFIKHQIESIQ 168

RESULT 25
US-08-470-896-107

; Sequence 107, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/470,896
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-107

APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,896
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-470-896-107

Query Match 14.4%; Score 412; DB 4; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;

Qy 318 VPILPFGVIGALGTGGITTTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
Db 2 IQILPFLVGLGTTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQDVSIAE 61

Qy 378 VVLQNRALDLTAERGTCFLGEECCYYNQSGIVTEKVEIRDRIRRAEELNTGP 437
Db 62 VVLQNRGLDLTAERGTCFLGEECCYYNQSGIVTEKVEIRDRIRRAEELNTGP 121

Qy 438 WGLLSQMPWILPPLGFLAIIILLFGPCIFNLLVNFVSSRIEAVK 484
Db 122 WTSFHGFLPYWPLGLPLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 24
US-08-485-546A-107
Sequence 107, Application US/08485546A
Patent No. 6518013
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 214

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,546A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-546A-107

Query Match 14.4%; Score 412; DB 4; Length 192;
Best Local Similarity 44.3%; Pred. No. 6e-34;
Matches 74; Conservative 42; Mismatches 51; Indels 0; Gaps 0;

Qy 318 VPILPFGVIGALGTGGITTTQFYKLSQELNGDMERVADSLVTLQDQNSLAA 377
Db 2 IQILPFLVGLGTTAVSTGAAGLGVSTQYTKLSHQLISDVQAISSTIQDQDVSIAE 61

Qy 378 VVLQNRALDLTAERGTCFLGEECCYYNQSGIVTEKVEIRDRIRRAEELNTGP 437
Db 62 VVLQNRGLDLTAERGTCFLGEECCYYNQSGIVTEKVEIRDRIRRAEELNTGP 121

Qy 438 WGLLSQMPWILPPLGFLAIIILLFGPCIFNLLVNFVSSRIEAVK 484
Db 122 WTSFHGFLPYWPLGLPLCLLLVLSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 25
US-09-376-781-6
Sequence 6, Application US/09376781
Patent No. 6261806
GENERAL INFORMATION:
APPLICANT: Banerjee, Papiya T.
APPLICANT: Pattence, Clive
APPLICANT: Andersson, Goran K.
TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of
USE
TITLE OF INVENTION: Use
FILE REFERENCE: 61750-267
CURRENT APPLICATION NUMBER: US/09/376,781
CURRENT FILING DATE: 1999-08-18
EARLIER APPLICATION NUMBER: 60/097,015
EARLIER FILING DATE: 1998-08-18
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 638
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: PERV-C
 ; OTHER INFORMATION: polypeptide sequence taken from GenBank Accession
 ; OTHER INFORMATION: No. 6261806 AF038600 for comparison.
 US-09-376-781-6

| | | | | |
|-----------------------|------------------|------------------|-------------|-------------|
| Query Match | 14.3% | Score 411.5; | DB 3; | Length 638; |
| Best Local Similarity | 24.7% | Pred. No. 5e-33; | | |
| Matches 136; | Conservative 76; | Mismatches 208; | Indels 131; | Gaps 19; |

| | | | | |
|----|-----|---|----------------------|-----|
| QY | 58 | GPPTTAHTHPRNCYHSATLCMHANTHYW----- | TGKMINSPCGGLGVTVCW-- | 106 |
| Db | 115 | GPNNKGKCGNPRDFCFCKQNCVTSNDGYKWPPTSQQDRVSFYVNTYTSSGQFNLTWI | 174 | |
| QY | 107 | -----TYFTOTGMSD----- | GGGVQDQAREKHVKEVIS | 134 |
| Db | 175 | RTGSPKCSFSDLYLKISFTEKQGENILKWNGNSGCMVYGGSGKQPGSILIRLKIN | 234 | |
| QY | 135 | QL-----TRVHTSSPY-KGLDLSKLHETLRTHRLVSLFNTLTGL | 175 | |
| Db | 235 | OLEPPMAIGPNTVLGTQRPPTQGGPSSNITSGSDPTSSSTTKMGAKLFSLIQGAFOAL | 294 | |
| QY | 176 | HEVSAQNPNCHICLPLNFRPVVISIPVPEQWNNFSTEINTTSVLGVLVSNLEITHSNL | 235 | |
| Db | 295 | NSTTPEATSSCWLCIASG-----PPYEGMARR----- | GKFNVTKEHRD | 333 |
| QY | 236 | TCV-----KFSNTTNTNSQCIRWTPPTQIVCLPESGIF----- | FVCGT | 274 |
| Db | 334 | QCTWGSQNKLTITEVSGKGTCLGKVPKPSHGLCNHTEAFNQTSQYLVPGYDRWWACNT | 393 | |
| QY | 275 | SAYRCIN-----GSSBSMCLFLVPPMTIYTEQ---DLYSYVISKPRNKRVPI--- | LPFV | 324 |
| Db | 394 | GLTPCVSTLVFNQTKDFCIMVQIVPRVYYPEKAILDEYD--RNHRQKREPISLTLMV | 451 | |
| QY | 325 | IGAGVLGALGTGIGGITTSTQ---FYKLSOELNGDMERVADSLVTLDQDQNSLAAYVL | 380 | |
| Db | 452 | LGLGVAAGVGTGAALVTGFPQQLGSLNHLRIVTEDLQALEKSVSNLEESLTSLEVVL | 511 | |
| QY | 381 | QNRALDLLTAERGGTCLFLGEECCVYNQSGIVTEKVKKEIDRIQRAAEELRNTGPW-- | 438 | |
| Db | 512 | QNRRLDILLFLKEGGLCVALKKECCCFYVDHSGAIRDSMNKLERLEKRRREKETTGWFE | 571 | |
| QY | 439 | GLL--SOMMEWILPPL-GPLAAIILLLFPGCIFNLLVNFVSSRIEAVKLOMEPRMQSKT | 495 | |
| Db | 572 | GWNRSLWLTLLSALTGTLVLLVLLLTGVCIIINKLIAFIRERISAVQI-MVLRQQ--- | 627 | |
| QY | 496 | KIYRPLDRPA | 506 | |
| Db | 628 | --YQSPSSREA | 636 | |

Search completed: April 19, 2004, 10:37:20
 Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:41:44 ; Search time 20 Seconds
(without alignments)
437.671 Million cell updates/sec

Title: US-10-069-883-1_COPY_448_538

Perfect score: 457

Sequence: 1 ILPLGLAAILLLPGPC.....PPEISAAQPLLRPNAGSS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 110 | 24.1 | 642 | 1 VCMVFG | env polyprotein - |
| 2 | 110 | 24.1 | 642 | 1 VCMVSG | env polyprotein pr |
| 3 | 110 | 24.1 | 642 | 1 T10533 | env polyprotein pr |
| 4 | 110 | 24.1 | 662 | 1 VCMVGF | env polyprotein - |
| 5 | 110 | 24.1 | 662 | 2 A25982 | env polyprotein - |
| 6 | 109 | 23.9 | 432 | 2 A25483 | env polyprotein, r |
| 7 | 108 | 23.6 | 563 | 1 VCMVM7 | env polyprotein - |
| 8 | 107 | 23.4 | 645 | 1 VCMVLS | env polyprotein pr |
| 9 | 107 | 23.4 | 662 | 1 A43491 | env polyprotein - |
| 10 | 107 | 23.4 | 688 | 2 B43491 | env polyprotein - |
| 11 | 107 | 23.4 | 689 | 2 A03982 | env polyprotein - |
| 12 | 105.5 | 23.1 | 77 | 2 A03982 | p15E protein - sim |
| 13 | 105.5 | 23.1 | 640 | 1 VCMVRF | env polyprotein pr |
| 14 | 105.5 | 23.1 | 676 | 1 VCMVVF | env polyprotein pr |
| 15 | 105.5 | 23.1 | 676 | 2 S70395 | env polyprotein - |
| 16 | 105.5 | 23.1 | 676 | 2 T01381 | env protein - mur |
| 17 | 105 | 23.0 | 636 | 1 VCMVFS | env polyprotein - |
| 18 | 105 | 23.0 | 644 | 2 S15464 | gp70 protein - mur |
| 19 | 105 | 23.0 | 661 | 1 VCMVCB | env polyprotein - |
| 20 | 105 | 23.0 | 668 | 1 VCMVFP | env polyprotein - |
| 21 | 104.5 | 22.9 | 627 | 1 VCMVM2 | env polyprotein - |
| 22 | 103 | 22.5 | 211 | 2 A03986 | env polyprotein - |
| 23 | 103 | 22.5 | 353 | 2 S22805 | env polyprotein - |
| 24 | 103 | 22.5 | 586 | 1 VCLJMP | env polyprotein - |
| 25 | 103 | 22.5 | 587 | 1 VCLJSA | env polyprotein - |
| 26 | 103 | 22.5 | 665 | 1 VCMWEM | env polyprotein - |
| 27 | 103 | 22.5 | 669 | 1 VCMWEK | env polyprotein - |
| 28 | 103 | 22.5 | 669 | 2 A46511 | envelope protein - |
| 29 | 102 | 22.3 | 201 | 2 A03985 | env polyprotein - |

| | | | | | |
|-----|------|------|------|----------|--------------------|
| 30 | 102 | 22.3 | 639 | 1 VCMVSA | env polyprotein pr |
| 31 | 102 | 22.3 | 666 | 1 VCMVHL | env polyprotein pr |
| 32 | 101 | 22.1 | 665 | 1 VCMVVR | env polyprotein pr |
| 33 | 101 | 22.1 | 665 | 1 VCMVKA | env polyprotein - |
| 34 | 100 | 21.9 | 321 | 2 S12812 | env polyprotein - |
| 35 | 97 | 21.2 | 567 | 1 VCFVAS | env polyprotein pr |
| 36 | 94 | 20.6 | 667 | 1 VCLJGL | retrovirus-related |
| 37 | 92 | 20.1 | 570 | 4 B44282 | env polyprotein - |
| 38 | 92 | 20.1 | 640 | 1 VCVMM1 | env polyprotein - |
| 39 | 88 | 19.3 | 582 | 1 VCVDA1 | env polyprotein - |
| 40 | 84 | 18.4 | 671 | 1 VCMVCE | proteinase homolog |
| 41 | 75 | 16.4 | 309 | 2 A83809 | probable transcrip |
| 42 | 71 | 15.5 | 445 | 2 T39415 | env polyprotein pr |
| 43 | 71 | 15.5 | 575 | 1 VCLJHD | hypothetical prote |
| 44 | 70 | 15.3 | 367 | 2 F71562 | hypothetical prote |
| 45 | 67 | 14.7 | 270 | 2 H83619 | amino acid transpo |
| 46 | 66.5 | 14.6 | 594 | 2 T43246 | hypothetical prote |
| 47 | 66 | 14.4 | 182 | 2 F70509 | tat binding protei |
| 48 | 66 | 14.4 | 448 | 2 T03776 | p83/100 protein - |
| 49 | 66 | 14.4 | 672 | 2 S61463 | p83/100 protein - |
| 50 | 66 | 14.4 | 693 | 2 S61464 | p83/100 protein - |
| 51 | 66 | 14.4 | 693 | 2 I40090 | p93 protein - lyme |
| 52 | 65.5 | 14.3 | 1165 | 2 T15279 | hypothetical prote |
| 53 | 65.5 | 14.3 | 2215 | 2 T16871 | hypothetical prote |
| 54 | 65 | 14.2 | 576 | 2 S61977 | transcription fact |
| 55 | 65 | 14.2 | 1178 | 2 S08405 | hypothetical prote |
| 56 | 65 | 14.2 | 6831 | 2 A88852 | protein unc-22 (im |
| 57 | 65 | 14.2 | 6839 | 2 S57242 | hypothetical prote |
| 58 | 65 | 14.2 | 7160 | 2 T27935 | hypothetical prote |
| 59 | 64.5 | 14.1 | 337 | 2 I38500 | interferon gamma r |
| 60 | 64.5 | 14.1 | 439 | 2 A12098 | hypothetical prote |
| 61 | 64.5 | 14.1 | 1432 | 2 B85431 | trichohyalin like |
| 62 | 64.5 | 14.1 | 1467 | 2 T23950 | hypothetical prote |
| 63 | 64 | 14.0 | 369 | 2 C82667 | probable methionin |
| 64 | 64 | 14.0 | 867 | 2 JH0225 | L96 protein - Tipu |
| 65 | 64 | 14.0 | 1206 | 2 S24407 | formin isoform IV |
| 66 | 63.5 | 13.9 | 51 | 2 T11263 | ATPase 8 - hardbac |
| 67 | 63.5 | 13.9 | 164 | 2 E70631 | AT103 protein - My |
| 68 | 63.5 | 13.9 | 524 | 2 T03112 | minor capsid scaff |
| 69 | 63.5 | 13.9 | 833 | 2 T01547 | probable phospholi |
| 70 | 63 | 13.8 | 747 | 2 AD2015 | hypothetical prote |
| 71 | 63 | 13.8 | 1257 | 2 T09493 | period protein hom |
| 72 | 63 | 13.8 | 1257 | 2 T13957 | period protein PER |
| 73 | 63 | 13.8 | 1268 | 1 A39640 | neural cell adhesi |
| 74 | 63 | 13.8 | 1541 | 2 T30227 | pipicolate-incorpo |
| 75 | 62.5 | 13.7 | 155 | 2 E82843 | hypothetical prote |
| 76 | 62.5 | 13.7 | 205 | 2 C75390 | NADH2 dehydrogenas |
| 77 | 62.5 | 13.7 | 385 | 2 B69114 | conserved hypotet |
| 78 | 62.5 | 13.7 | 574 | 2 B88465 | protein B0244.8 li |
| 79 | 62.5 | 13.7 | 681 | 2 T00905 | hypothetical prote |
| 80 | 62.5 | 13.7 | 770 | 2 T15742 | hypothetical prote |
| 81 | 62.5 | 13.7 | 854 | 2 S02003 | neurofilament trip |
| 82 | 62.5 | 13.7 | 1072 | 1 A37221 | neurofilament trip |
| 83 | 62.5 | 13.7 | 1784 | 2 T43167 | sodium channel pro |
| 84 | 62 | 13.6 | 309 | 2 G69379 | proteinase homolog |
| 85 | 62 | 13.6 | 354 | 2 T31780 | hypothetical prote |
| 86 | 62 | 13.6 | 438 | 2 A83037 | probable MFS trans |
| 87 | 62 | 13.6 | 611 | 2 T04510 | hypothetical prote |
| 88 | 62 | 13.6 | 1277 | 2 T30532 | neural cell adhesi |
| 89 | 62 | 13.6 | 2064 | 2 T13707 | still life protein |
| 90 | 61.5 | 13.5 | 111 | 1 A32233 | thioredoxin 2 (val |
| 91 | 61.5 | 13.5 | 115 | 2 A82898 | probable membrane |
| 92 | 61.5 | 13.5 | 206 | 2 T34561 | hypothetical prote |
| 93 | 61.5 | 13.5 | 434 | 2 H83227 | cation efflux syst |
| 94 | 61.5 | 13.5 | 1075 | 2 S76433 | hypothetical prote |
| 95 | 61 | 13.3 | 205 | 2 E72367 | ct345 hypothetical |
| 96 | 61 | 13.3 | 288 | 2 A72128 | hypothetical prote |
| 97 | 61 | 13.3 | 288 | 2 D81543 | hypothetical prote |
| 98 | 61 | 13.3 | 288 | 2 D86494 | CT345 hypothetical |
| 99 | 61 | 13.3 | 721 | 2 A33319 | microtubule-associ |
| 100 | 60.5 | 13.2 | 238 | 2 T29250 | hypothetical prote |

ALIGNMENTS

RESULT 1

VCMVFG
 env polyprotein - feline leukemia virus (strain A/Glasgow-1)
 N;Alternate names: coat polyprotein
 N;Contains: coat protein p15E; knob protein gp70
 C;Species: feline leukemia virus
 C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 12-Apr-1996
 C;Accession: A24300
 R;Stewart, M.A.; Warnock, M.; Wheeler, A.; Wilkie, N.; Mullins, J.I.; Onions, D.E.; Neil
 J. Virol. 58, 825-834, 1986
 A;Title: Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and lo
 A;Reference number: A24300; MUID:86200439; PMID:3009890
 A;Accession: A24300
 A;Molecule type: DNA
 A;Residues: 1-642 <STE>
 A;Note: the authors translated the codon GCT for residue 158 as Thr and ACC for residue

C;Genetics:
 C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
 F;1-33/Domain: signal sequence #status predicted <SIG>
 F;34-445/Product: knob protein gp70 #status predicted <GP7>
 F;446-642/Product: coat protein p15E #status predicted <P15>
 F;35,43,58,91,267,302,307,334,374,390,410,578/Binding site: carbohydrate (Asn) (covalent

Query Match 24.1%; Score 110; DB 1; Length 642;
 Best Local Similarity 43.1%; Pred. No. 0.00074;
 Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAIILLILFGPCIFNLLVNFVSSRIEAVKLMQEPKMQSKTKIYRRPLDRP 58

Db 587 ISSIMGPELLILLILFGPCILNRLVQFVKDRISVVQALILTQQYQIQKY--DPDRP 642

RESULT 2

VCMVS2
 env polyprotein precursor - feline sarcoma virus (strain GA)
 N;Contains: coat protein gp70; coat protein p15E
 C;Species: feline sarcoma virus
 A;Note: host Felis silvestris catus (domestic cat)
 C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
 C;Accession: B33741
 R;Guillot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
 Virology 161, 252-258, 1987
 A;Title: Nucleotide sequence analysis of the LTRs and env genes of SM-FesV and GA-FesV.

A;Reference number: A33741; MUID:88044502; PMID:2823466
 A;Accession: B33741
 A;Molecule type: DNA
 A;Residues: 1-642 <GUI>
 C;Genetics:

A;Gene: env
 C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
 F;1-32/Domain: signal sequence #status predicted <SIG>
 F;33-445/Product: coat protein gp70 #status predicted <GP>
 F;446-642/Product: coat protein p15E #status predicted <GUO>
 F;34,42,57,267,302,307,331,334,374,390,410,578/Binding site: carbohydrate (Asn) (covalent

Query Match 24.1%; Score 110; DB 1; Length 642;
 Best Local Similarity 43.1%; Pred. No. 0.00074;
 Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAIILLILFGPCIFNLLVNFVSSRIEAVKLMQEPKMQSKTKIYRRPLDRP 58

Db 587 ISSIMGPELLILLILFGPCILNRLVQFVKDRISVVQALILTQQYQIQKY--DPDRP 642

RESULT 3

T10533
 env polyprotein precursor - feline leukemia virus (strain Felv-FAIDS)

N;Contains: env protein gp70; env protein p15E
 C;Species: feline leukemia virus
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C;Accession: T10533
 R;Donahue, P.R.; Hoover, E.A.; Beltz, G.A.; Riedel, N.; Hirsch, V.M.; Overbaugh, J.; Mul
 J. Virol. 62, 722-731, 1988
 A;Title: Strong sequence conservation among horizontally transmissible, minimally pathog

A;Reference number: Z17078; MUID:88119207; PMID:2828667
 A;Accession: T10533
 A;Status: preliminary; translated from GE/EMBL/DBD

A;Molecule type: mRNA
 A;Residues: 1-642 <DON>
 A;Cross-references: EMBL:M18247; NID:G323904; PIDN:AAA93093.1; PID:G323906
 C;Superfamily: type C retrovirus env polyprotein

Query Match 24.1%; Score 110; DB 2; Length 642;
 Best Local Similarity 43.1%; Pred. No. 0.00074;
 Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAIILLILFGPCIFNLLVNFVSSRIEAVKLMQEPKMQSKTKIYRRPLDRP 58

Db 587 ISSIMGPELLILLILFGPCILNRLVQFVKDRISVVQALILTQQYQIQKY--DPDRP 642

RESULT 4

VCMVGF
 env polyprotein - feline leukemia virus (strain Gardner-Arnstein)
 N;Contains: knob protein gp70; R protein; spike protein p15E
 C;Species: feline leukemia virus
 A;Note: host Felis sp. (cat)
 C;Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999
 C;Accession: A03991
 R;Nunberg, J.H.; Williams, M.E.; Innis, M.A.

J. Virol. 49, 629-632, 1984
 A;Title: Nucleotide sequences of the envelope genes to two isolates of feline leukemia v
 A;Reference number: A92996; MUID:84115095; PMID:6319767
 A;Accession: A03991

A;Molecule type: DNA
 A;Residues: 1-662 <NUN>

A;Cross-references: GB:X01209; NID:G323910; PIDN:AAA43052.1; PID:G323911
 R;Elder, J.H.; McGe, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Gran
 J. Virol. 61, 8-15, 1987

A;Title: Localization of neutralizing regions of the envelope gene of feline leukemia vi
 A;Reference number: A25982; MUID:87061257; PMID:2431166

A;Contents: annotation; peptide synthesis
 A;Note: 27 peptides synthesized

C;Genetics:
 C;Gene: env

C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei

F;1-33/Domain: signal sequence #status predicted <SIG>
 F;34-465/Product: knob protein gp70 #status predicted <KPG>
 F;466-645/Product: spike protein p15E #status predicted <SKP>
 F;646-662/Product: R protein #status predicted <RRP>

F;43,58,286,322,327,351,354,394,410,430/Binding site: carbohydrate (Asn) (covalent) #sta

Query Match 24.1%; Score 110; DB 1; Length 662;
 Best Local Similarity 43.1%; Pred. No. 0.00076;
 Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAIILLILFGPCIFNLLVNFVSSRIEAVKLMQEPKMQSKTKIYRRPLDRP 58

Db 607 ISSIMGPELLILLILFGPCILNRLVQFVKDRISVVQALILTQQYQIQKY--DPDRP 662

RESULT 5

A25982
 env polyprotein - feline leukemia virus (strain Gardner-Arnstein)
 C;Species: feline leukemia virus
 C;Date: 25-Oct-1987 #sequence_revision 25-Oct-1987 #text_change 24-Feb-1995

C;Accession: A25982
 R;Elder, J.H.; McGe, J.S.; Munson, M.; Houghten, R.A.; Kloetzer, W.; Bittle, J.L.; Gran

J. Virol. 61, 8-15, 1987

A:Title: Localization of neutralizing regions of the envelope gene of feline leukemia virus
A:Reference number: A25982; UID:87061257; PMID:2431166
A:Accession: A25982
A:Molecule type: protein
A:Residues: 1-662 <ELD>
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein

Query Match 24.1%; Score 110; DB 2; Length 662;
Best Local Similarity 43.1%; Pred. No. 0.00075;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAILLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 58
DB 607 ISSMGPLVILLLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 662

RESULT 6
VCNWB
env polypeptide precursor - mouse (fragments)
N:Contains: coat protein gp70; coat protein p15E
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 20-Sep-1999
C:Accession: A25483; B25483
R:Ikeda, H.; Laigret, F.; Martin, M.A.; Repaske, R.
J. Virol. 55, 768-777, 1985
A:Title: Characterization of a molecularly cloned retroviral sequence associated with FV-1
A:Reference number: A25483; UID:85265041; PMID:2991595
A:Accession: A25483
A:Molecule type: DNA
A:Residues: 1-432 <IKE>
A:Accession: B25483
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 349-432 <IK2>
A:CROSS-references: GB:M11052; NID:G193055; PIDN:AAA37563.1; PID:G387148
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein

Query Match 23.9%; Score 109; DB 2; Length 432;
Best Local Similarity 32.5%; Pred. No. 0.00061;
Matches 26; Conservative 12; Mismatches 22; Indels 20; Gaps 2;

QY 1 ILPLGLPLAAILLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 60
DB 373 ISTMGPLVILLLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 417

QY 61 PRSDVNDIKGTPPEISAAQ 80
DB 418 -----HQLKSDPEVESRE 432

RESULT 7
VCNWB
env polypeptide - baboon endogenous virus (strain M7)
N:Alternate names: coat polypeptide
N:Contains: coat protein gp70; coat protein p20E
C:Species: baboon endogenous virus
A:Note: host Papio sp. (baboon)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
C:Accession: JTC0262
R:Kato, S.; Matsuo, K.; Nishimura, N.; Takahashi, N.; Takano, T.
Jpn. J. Genet. 62, 127-137, 1987
A:Title: The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome
A:Reference number: JTC0262
A:Accession: JTC0262
A:Molecule type: DNA
A:Residues: 1-563 <KAT>
A:CROSS-references: GB:M16550; NID:G509586; PIDN:AAA87333.1; PID:G332599
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide

C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F;1-376/Product: coat protein gp70 #status predicted <CGP>
F;377-563/Product: coat protein p20E #status predicted <CGP>
F;437-449/Region: immunosuppressive peptide #status predicted
F;113,219,229,264,282,292,306,312,321,339,469/Binding site: carbohydrate (Asn) (covalent)

Query Match 23.6%; Score 108; DB 1; Length 563;
Best Local Similarity 50.0%; Pred. No. 0.001;
Matches 18; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAILLLGPGCIFNLLVNFVSSRIEAV 36
DB 508 LLPFLGPLLTLTLLTGTCIFNLLVNFVSSRIEAV 543

RESULT 8
VCNWB
env polypeptide precursor - feline sarcoma virus (strain SM)
N:Contains: coat protein gp70; coat protein p15E
C:Species: feline sarcoma virus
A:Note: host Felis silvestris catus (domestic cat)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 12-Apr-1996
C:Accession: A33741
R:Guinhot, S.; Hampe, A.; D'Auriol, L.; Galibert, F.
Virology 161, 252-258, 1987
A:Title: Nucleotide sequence analysis of the LTRs and env genes of SM-FesV and GA-FesV.
A:Reference number: A33741; UID:88044502; PMID:2823466
A:Accession: A33741
A:Molecule type: DNA
A:Residues: 1-645 <GUI>
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F;1-36/Domain: signal sequence #status predicted <SIG>
F;37-448/Product: coat protein gp70 #status predicted <GUP>
F;449-645/Product: coat protein p15E #status predicted <GUO>
F;38,46,61,270,305,310,334,337,377,393,413,581/Binding site: carbohydrate (Asn) (covalent)

Query Match 23.4%; Score 107; DB 1; Length 645;
Best Local Similarity 43.1%; Pred. No. 0.0016;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGLPLAAILLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 58
DB 590 ISSMGPLVILLLLGPGCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRPLDRP 645

RESULT 9
VCNWB
env polypeptide - feline leukemia virus (strain lambda-B1) (fragment)
N:Contains: coat protein gp70; coat protein p15E
C:Species: feline leukemia virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C:Accession: A27172
R:Nicolaisen-Strouss, K.; Kumar, H.P.M.; Fitting, T.; Grant, C.K.; Eider, J.H.
J. Virol. 61, 3410-3415, 1987
A:Title: Natural feline leukemia virus variant escapes neutralization by a monoclonal antibody
A:Reference number: A27172; UID:88036192; PMID:2444714
A:Accession: A27172
A:Molecule type: DNA
A:Residues: 1-662 <NIC>
A:CROSS-references: GB:J03448; NID:G323896; PIDN:AAA43048.1; PID:G323897
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F;1-33/Domain: signal sequence #status predicted <SIG>
F;34-465/Product: coat protein gp70 #status predicted <GPS>
F;466-662/Product: coat protein p15E #status predicted <GPT>
F;43,58,286,322,327,351,354,430/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.4%; Score 107; DB 1; Length 662;

Best Local Similarity 43.1%; Pred. No. 0.0016;
Matches 22; Conservative 8; Mismatches 21; Indels 0; Gaps 0;

QY 1 ILPFLGLAAIILLGLFGPCIFNLNVFSSRIEAVKLMQEPKMSKTKIY 51
Db 607 ISTIMGPLILLILLGLFGPCILNRLVQFVKDRISVVQALILTQQYQKQY 657

RESULT 10
A43491
env polyprotein - T3651/B murine leukemia virus (strain T1223/B)
C:Species: T3651/B murine leukemia virus
C>Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 20-Apr-2001
C:Accession: A43491
R:Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand, Virus Res. 18, 117-134, 1991
A:Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-in
A:Reference number: A43491; MUID:91253260; PMID:2042396
A:Accession: A43491
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-688 <BAY>
A:Cross-references: GB:X59002
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 23.4%; Score 107; DB 2; Length 688;
Best Local Similarity 31.2%; Pred. No. 0.0017;
Matches 25; Conservative 13; Mismatches 22; Indels 20; Gaps 2;

QY 1 ILPFLGLAAIILLGLFGPCIFNLNVFSSRIEAVKLMQEPKMSKTKIYRPLDRPAS 60
Db 629 ISTIMGPLILLILLGLFGPCILNRLVQFVKDRISVVQALV-----LQQY----- 673

QY 61 PRSDVNDIKGTPPEISAAQ 80
Db 674 -----HQLKSIDPEEVESRE 688

RESULT 11
B43491
env polyprotein - T3651/B murine leukemia virus
C:Species: T3651/B murine leukemia virus
C>Date: 12-Jan-1993 #sequence_revision 12-Jan-1993 #text_change 20-Apr-2001
C:Accession: B43491
R:Baylac-Kalabokias, H.; Astier-Gin, T.; Moynet, D.; Hernould, M.; Mamoun, R.; Legrand, Virus Res. 18, 117-134, 1991
A:Title: A new leukemogenic retrovirus isolated from tumor cells derived from a radio-in
A:Reference number: A43491; MUID:91253260; PMID:2042396
A:Accession: B43491
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-689 <BAY>
A:Cross-references: GB:X59001
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 23.4%; Score 107; DB 2; Length 689;
Best Local Similarity 31.2%; Pred. No. 0.0017;
Matches 25; Conservative 13; Mismatches 22; Indels 20; Gaps 2;

QY 1 ILPFLGLAAIILLGLFGPCIFNLNVFSSRIEAVKLMQEPKMSKTKIYRPLDRPAS 60
Db 630 ISTIMGPLILLILLGLFGPCILNRLVQFVKDRISVVQALV-----LQQY----- 674

QY 61 PRSDVNDIKGTPPEISAAQ 80
Db 675 -----HQLKSIDPEEVESRE 689

RESULT 12
A03982
p15E protein - simian sarcoma virus (fragment)

C:Species: simian sarcoma virus
C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 20-Jun-2000
C:Accession: A03982
R:Devare, S.G.; Reddy, E.P.; Law, J.D.; Robbins, K.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 80, 731-735, 1983
A:Title: Nucleotide sequence of the simian sarcoma virus genome: demonstration that its
A:Reference number: A03982; MUID:93144004; PMID:6298772
A:Accession: A03982
A:Molecule type: genomic RNA
A:Residues: 1-77 <DEV>
A:Cross-references: GB:V01201; GB:J02394; GB:J02396; GB:J02397; NID:961774; PIDN:CAA2451
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein

Query Match 23.1%; Score 105.5; DB 2; Length 77;
Best Local Similarity 50.0%; Pred. No. 0.00021;
Matches 21; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

QY 6 GPLAAIILLGLFGPCIFNLNVFSSRIEAVK-LQMEPKMS 46
Db 29 GPLILLILLGLFGPCILNRLVQFINNRVSAVKILVLRQKYQT 70

RESULT 13
VCMVRV
env polyprotein precursor - Rauscher mink cell focus-forming virus
N:Contains: coat protein p12E; coat protein p15E; knob protein gp70
C:Species: Rauscher mink cell focus-forming virus
C>Date: 30-Jun-1987 #sequence_revision 30-Jun-1987 #text_change 16-Jul-1999
C:Accession: A03990
R:Vogt, M.; Hagglom, C.; Swift, S.; Haas, M.
J. Virol. 55, 184-192, 1985
A:Title: Envelope gene and long terminal repeat determine the different biological proper
A:Reference number: A93011; MUID:85237696; PMID:4009793
A:Accession: A03990
A:Molecule type: DNA
A:Residues: 1-640 <VOG>
A:Cross-references: GB:M10100; NID:G332068; PIDN:AAA6528.1; PID:G332070
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polyprotein
C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protei
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-443/Product: knob protein gp70 #status predicted <KBP>
F:444-640/Product: coat protein p15E #status predicted <PTE>
F:444-623/Product: coat protein p12E #status predicted <PTE>
F:43,58,300,332,339,372,408,576/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 23.1%; Score 105.5; DB 1; Length 640;
Best Local Similarity 41.1%; Pred. No. 0.0022;
Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;

QY 1 ILPFLGLAAIILLGLFGPCIFNLNVFSSRIEAVK-----LQMEPK 43
Db 585 ISTIMGPLILLILLGLFGPCILNRLVQFVKDRISVVQALVLTQQYHQLKPEVEPQ 640

RESULT 14
VCMVRV
env polyprotein precursor - Friend murine leukemia virus (strain PVC-211)
N:Alternate names: coat polyprotein
N:Contains: knob protein gp76; R protein; spike protein p15E
C:Species: Friend murine leukemia virus
C>Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 16-Jul-1999
C:Accession: A38210; S35476
R:Masuda, M.; Remington, M.P.; Hoffman, P.M.; Ruscetti, S.K.
J. Virol. 66, 2798-2806, 1992
A:Title: Molecular characterization of a neuropathogenic and nonerythroleukemogenic vari
A:Reference number: A38210; MUID:92219364; PMID:1560524
A:Accession: A38210
A:Molecule type: DNA
A:Residues: 1-676 <SPE>

A;Cross-references: GB:M93134; NID:G331898; PIDN:AAA46478.1; PID:G331901
 R;Remington, M.P.; Hoffman, P.M.; Ruscelletti, S.K.; Masuda, M.
 Nucleic Acids Res. 20, 3249, 1992
 A;Title: Complete nucleotide sequence of a neuropathogenic variant of Friend murine leukemia virus
 A;Reference number: S35474; MUID:92319660; PMID:1620621
 A;Accession: S35476
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-676 <REM>
 A;Cross-references: EMBL:M93134; NID:G331898; PIDN:AAA46478.1; PID:G331901
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1992
 C;Genetics:
 A;Gene: env
 C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
 F;1-35/Domain: signal sequence #status predicted <SIG>
 F;36-676/Product: env polyprotein #status predicted <ENV>
 F;36-619/Domain: extracellular #status predicted <EXT>
 F;36-479/Product: knob protein gp76 #status predicted <KGP>
 F;476-479/Region: cleavage processing #status predicted
 F;480-659/Product: spike protein p15E #status predicted <SPP>
 F;486-502/Region: hydrophobic #status predicted
 F;620-636/Domain: transmembrane #status predicted <TM1>
 F;637-676/Domain: intracellular #status predicted <INT>
 F;660-676/Product: R protein #status predicted <RPT>
 F;46,202,336,368,408,444/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.1%; Score 105.5; DB 1; Length 676;
 Best Local Similarity 41.1%; Pred. No. 0.0024;
 Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVK-----LQMEPK 43
 Db 621 ISTIMGPLIILLLLFGPCILNRLVQFVKDRISVVQALVLTQQYHQLKPLEYEPQ 676

RESULT 15
 S70395
 env polyprotein - Friend murine leukemia virus (strain FB29)
 N;Contains: knob protein gp76; R protein; spike protein p15E
 C;Species: Friend murine leukemia virus
 A;Variety: strain FB29
 C;Accession: S70395
 C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
 R;Perryman, S.; Nishio, J.; Chesebro, B.
 Nucleic Acids Res. 19, 6950, 1991
 A;Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.
 A;Reference number: S70393; MUID:92107687; PMID:1762923
 A;Accession: S70395
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: genomic RNA
 A;Residues: 1-676 <PER>
 A;Cross-references: EMBL:211128; NID:G61547; PIDN:CAA77479.1; PID:G61549
 A;Experimental source: strain FB29
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
 C;Genetics:
 A;Gene: env
 C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; polyprotein; spike protein; transmembrane protein
 Query Match 23.1%; Score 105.5; DB 2; Length 676;
 Best Local Similarity 41.1%; Pred. No. 0.0024;
 Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVK-----LQMEPK 43
 Db 621 ISTIMGPLIILLLLFGPCILNRLVQFVKDRISVVQALVLTQQYHQLKPLEYEPQ 676

RESULT 16
 T01381
 env protein - murine leukemia virus
 C;Species: murine leukemia virus

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Jun-2000
 C;Accession: T01381
 R;Tamas, K.M.; Pozsgay, J.M.; Avidan, N.; Ksiazek, S.J.; Overmoyer, B.; Blank, K.J.; Pry
 Virology 192, 587-595, 1993
 A;Title: Loss of antigenic epitopes as the result of env gene recombination in retrovirus
 A;Reference number: Z14313; MUID:93134803; PMID:7678475
 A;Accession: T01381
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-676 <TUM>
 A;Cross-references: EMBL:S53043; NID:G263651
 C;Genetics:
 A;Gene: env
 C;Superfamily: type C retrovirus env polyprotein
 Query Match 23.1%; Score 105.5; DB 2; Length 676;
 Best Local Similarity 41.1%; Pred. No. 0.0024;
 Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVK-----LQMEPK 43
 Db 621 ISTIMGPLIILLLLFGPCILNRLVQFVKDRISVVQALVLTQQYHQLKPLEYEPQ 676

RESULT 17
 VCVWFS
 env polyprotein - mink cell focus-forming virus
 N;Alternate names: coat polyprotein
 N;Contains: knob protein gp70; R protein; spike protein p15E
 C;Species: mink cell focus-forming virus
 C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 23-Aug-1997
 C;Accession: A31668
 R;Chattopadhyay, S.K.; Baroudy, B.M.; Holmes, K.L.; Fredrickson, T.N.; Lander, M.R.; Mor
 Virology 168, 90-100, 1989
 A;Title: Biologic and molecular genetic characteristics of a unique MCF virus that is hi
 A;Reference number: A31668; MUID:89085614; PMID:2535909
 A;Accession: A31668
 A;Molecule type: DNA
 A;Residues: 1-636 <CHA>
 C;Genetics:
 A;Gene: env
 C;Superfamily: type C retrovirus env polyprotein
 C;Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein
 F;1-30/Domain: signal sequence #status predicted <SIG>
 F;31-440/Product: knob protein gp70 #status predicted <GP7>
 F;441-620/Product: spike protein p15E #status predicted <P15>
 F;621-636/Product: R protein #status predicted <RPT>
 F;43,58,297,329,369,405/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.0%; Score 105; DB 1; Length 636;
 Best Local Similarity 54.1%; Pred. No. 0.0025;
 Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVK 37
 Db 582 ISTIMGPLIILLLLFGPCILNRLVQFVKDRISVVQ 618

RESULT 18
 S15464
 gp70 protein - murine leukemia virus
 C;Species: murine leukemia virus
 C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Sep-1999
 C;Accession: S15464; S15465
 R;Maruyama, N.
 submitted to the EMBL Data Library, April 1991
 A;Reference number: S15464
 A;Accession: S15464
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-644 <VAR>
 A;Cross-references: EMBL:X59305; NID:G59254; PIDN:CAA41992.1; PID:G59255; EMBL:X59304; N
 C;Superfamily: type C retrovirus env polyprotein

Query Match 23.0%; Score 105; DB 2; Length 644;
Best Local Similarity 54.1%; Pred. No. 0.0025;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 585 ISTINGPLIIVLLILLFGPCILNRLVQFVKDRISVVQ 621

RESULT 19
VCWVCP
env polypeptide - Cas-Br-E murine leukemia virus
N:Contains: coat protein p15E; knob protein gp70 precursor
C:Species: Cas-Br-E murine leukemia virus
C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jul-1999
C:Accession: B26103
R:Rassart, E.; Nelbach, L.; Jolicœur, P.
J. Virol. 60, 910-919, 1986
A:Title: Cas-Br-E murine leukemia virus: sequencing of the paralytogenic region of its genome
kemic tissues.
A:Reference number: A26103; MUID:87061215; PMID:3023680
A:Accession: B26103
A:Molecule type: DNA
A:Residues: 1-661 <RAS>
A:Cross-references: GB:M14702; NID:G332016; PIDN:AAA46512.1; PID:G332018
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-465/Product: knob protein gp70 #status predicted <KOP>
F:466-661/Product: coat protein p15E #status predicted <PTE>
F:43, 186, 199, 322, 361, 394, 430, 598/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.0%; Score 105; DB 1; Length 661;
Best Local Similarity 54.1%; Pred. No. 0.0026;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 607 ISTINGPLIIVLLILLFGPCILNRLVQFVKDRISVVQ 643

RESULT 20
VCWVFP
env polypeptide precursor (clone CFE-6) - feline leukemia virus (provirus)
N:Contains: coat protein gp70; coat protein p15E
C:Species: feline leukemia virus
A:Note: host Felis silvestris catus (domestic cat)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A31479
R:Kumar, D.V.; Berry, B.T.; Roy-Burman, P.
J. Virol. 63, 2379-2384, 1989
A:Title: Nucleotide sequence and distinctive characteristics of the env gene of endogenous feline leukemia virus
A:Reference number: A31479; MUID:8919802; PMID:2539525
A:Accession: A31479
A:Molecule type: DNA
A:Residues: 1-668 <KUM>
A:Cross-references: GB:M25425; NID:G163849; PIDN:AAA30809.1; PID:G163850
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-465/Product: coat protein gp70 #status predicted <GUP>
F:466-668/Product: coat protein p15E #status predicted <PPP>
F:43, 58, 286, 322, 327, 351, 354, 394, 410, 430, 597/Binding site: carbohydrate (Asn) (covalent)

Query Match 23.0%; Score 105; DB 1; Length 668;
Best Local Similarity 39.4%; Pred. No. 0.0026;
Matches 26; Conservative 11; Mismatches 25; Indels 4; Gaps 2;

QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 606 ISSINGPLIIVLLILLFGPCILNRLVQFVKDRISVVQ 662

RESULT 21
VCWVM2
env polypeptide - Friend mink cell focus-forming virus
N:Contains: knob protein gp70; spike protein p15E
C:Species: Friend mink cell focus-forming virus
C>Date: 20-Sep-1984 #sequence_revision 20-Sep-1984 #text_change 16-Jul-1999
C:Accession: A03989
R:Koch, W.; Zimmermann, W.; Oliff, A.; Friedrich, R.
J. Virol. 49, 828-840, 1984
A:Title: Molecular analysis of the envelope gene and long terminal repeat of Friend mink cell focus-forming virus
A:Reference number: A03989; MUID:84138778; PMID:6321768
A:Accession: A03989
A:Molecule type: DNA
A:Residues: 1-627 <KOC>
A:Cross-references: GB:M12528; NID:G331918; PIDN:AAA46483.1; PID:G331920
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-431/Product: knob protein gp70 #status predicted <KNS>
F:432-627/Product: spike protein p15E #status predicted <SPK>
F:43, 58, 288, 320, 360, 364, 396/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 22.9%; Score 104.5; DB 1; Length 627;
Best Local Similarity 41.8%; Pred. No. 0.0028;
Matches 23; Conservative 6; Mismatches 13; Indels 13; Gaps 1;

QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVK 42
DB 573 ISTINGPLIIVLLILLFGPCILNRLVQFVKDRISVVQ 627

RESULT 22
A03986
env polypeptide - FBJ murine osteosarcoma virus (fragment)
N:Contains: knob protein gp70; R protein; spike protein p15E
C:Species: FBJ murine osteosarcoma virus
C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 20-Jun-2000
C:Accession: A03986
R:Van Beveren, C.; van Straaten, F.; Curran, T.; Mueller, R.; Verma, I.M.
Cell 32, 1241-1255, 1983
A:Title: Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular proteins interact to enhance c-fos transcription
A:Reference number: A03986; MUID:93180421; PMID:6301687
A:Accession: A03986
A:Molecule type: DNA
A:Residues: 1-211 <VAN>
A:Cross-references: GB:V01184; NID:G61642; PIDN:CAA24506.1; PID:G1263006
A:Note: specific enzymatic cleavages in vivo yield mature proteins including the carboxy terminal domain
C:Genetics:
A:Gene: env
C:Superfamily: type C retrovirus env polypeptide
C:Keywords: coat protein; glycoprotein; polypeptide; spike protein; transmembrane protein
Query Match 22.5%; Score 103; DB 2; Length 211;
Best Local Similarity 51.4%; Pred. No. 0.0012;
Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 152 ISTINGPLIIVLLILLFGPCILNRLVQFVKDRISVVQ 188

RESULT 23

S22805
 env polyprotein - murine endogenous virus ENV-1 (fragment)
 C:Species: murine endogenous virus ENV-1
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
 C:Accession: S22805
 R:Timmons, P.M.; Brickell, P.M.; Latchman, D.S.; Rigby, P.W.J.
 Nucleic Acids Res. 19, 7215-7218, 1991
 A:Title: Activation of endogenous retroviral transcription in SV40-transformed mouse cells
 A:Reference number: S22805; MUID:92115572; PMID:1662812
 A:Accession: S22805
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-353 <TIM>
 A:Cross-references: EMBL:X62670; NID:G50850; PIDN:CAA44543.2; PID:G5701882
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
 C:Genetics:
 A:Gene: env
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: coat protein; glycoprotein; polyprotein; spike protein; transmembrane protein

Query Match 22.5%; Score 103; DB 2; Length 353;
 Best Local Similarity 51.4%; Pred. No. 0.0021;
 Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILPLGPIAAIILLLFGPCIFNLLVNFVSSRIEAVK 37
 DB 296 ISTMGPIILLLLLFGPCILNRLVQFIKDRISVQ 332

RESULT 24
 VCLJMP
 env polyprotein - Mason-Pfizer monkey virus
 N:Alternate names: coat polyprotein
 N:Contains: coat protein gp20; coat protein gp70
 C:Species: Mason-Pfizer monkey virus
 C>Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 16-Jul-1999
 C:Accession: D25839
 R:Sonigo, P.; Barker, C.; Hunter, E.; Wain-Hobson, S.
 Cell 45, 375-385, 1986
 A:Title: Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type virus
 A:Reference number: A90878; MUID:86189951; PMID:2421920
 A:Accession: D25839
 A:Molecule type: DNA
 A:Residues: 1-586 <SON>
 A:Cross-references: GB:M12349; NID:G334702; PIDN:AAA47712.1; PID:G334705
 A:Experimental source: clone 6A
 C:Genetics:
 A:Gene: env
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: capsid protein; coat protein; glycoprotein; polyprotein
 F:1-20/Domain: signal sequence #status predicted <SIG>
 F:21-394/Product: coat protein gp70 #status predicted <GP7>
 F:395-586/Product: coat protein gp20 #status predicted <GP2>
 F:120,237,264,276,291,304,318,324,339,357,487/Binding site: carbohydrate (Asn) (covalent)

Query Match 22.5%; Score 103; DB 1; Length 586;
 Best Local Similarity 45.9%; Pred. No. 0.0037;
 Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLGPIAAIILLLFGPCIFNLLVNFVSSRIEAVK 37
 DB 526 VMPGLGPIILLLLLFGPFIIFNKLMTFKQIESIQ 562

RESULT 25
 VCLJSA
 env polyprotein - simian AIDS retrovirus SRV-1
 N:Alternate names: coat polyprotein
 N:Contains: coat protein gp20; coat protein gp70
 C:Species: simian AIDS retrovirus SRV-1
 A:Note: host Macaca mulatta (rhesus macaque)
 C>Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 16-Jul-1999
 C:Accession: A04000

R:Power, M.D.; Marx, P.A.; Bryant, M.L.; Gardner, M.B.; Barr, P.J.; Luciw, P.A.
 Science 231, 1567-1572, 1986
 A:Title: Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome virus
 A:Reference number: A94711; MUID:86151668; PMID:3006247
 A:Accession: A04000
 A:Molecule type: DNA
 A:Residues: 1-587 <POW>
 A:Cross-references: GB:M11841; NID:G334746; PIDN:AAA47733.1; PID:G334750
 C:Comment: The env polyprotein contains coat protein gp70 and coat protein gp20; however
 C:Genetics:
 A:Gene: env
 C:Superfamily: type C retrovirus env polyprotein
 C:Keywords: capsid protein; coat protein; glycoprotein; polyprotein
 F:120,237,266,271,277,280,295,308,322,328,340,358,488,585/Binding site: carbohydrate (Asn)

Query Match 22.5%; Score 103; DB 1; Length 587;
 Best Local Similarity 45.9%; Pred. No. 0.0038;
 Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLGPIAAIILLLFGPCIFNLLVNFVSSRIEAVK 37
 DB 527 VMPGLGPIILLLLLFGPFIIFNKLMTFKQIESIQ 563

Search completed: April 19, 2004, 10:55:00
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:37:59 ; Search time 14 Seconds
(without alignments)

338.456 Million cell updates/sec

Title: US-10-069-883-1_COPY_448_538

Perfect score: 457

Sequence: 1 ILPFLGFLAAIILLFLGPGC.....PPEEISAAQPLLPNSAGSS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 110 | 24.1 | 642 | 1 ENV FLVGL | P08359 feline leuk |
| 2 | 110 | 24.1 | 662 | 1 ENV FSVGA | P03391 feline sarc |
| 3 | 109 | 23.9 | 679 | 1 ENV2 MOUSE | P11370 mus musculus |
| 4 | 108 | 23.6 | 563 | 1 ENV BAEVM | P10269 baboon endo |
| 5 | 107 | 23.4 | 574 | 1 ENV SRV2 | P51515 simian retr |
| 6 | 107 | 23.4 | 574 | 1 ENV FSVM | P21445 feline sarc |
| 7 | 107 | 23.4 | 662 | 1 ENV FLVLB | P11261 feline leuk |
| 8 | 105.5 | 23.1 | 77 | 1 ENV SMSAV | P03384 simian sarc |
| 9 | 105.5 | 23.1 | 640 | 1 ENV RMCVF | P06445 rauscher mi |
| 10 | 105.5 | 23.1 | 676 | 1 ENV MLVFF | P26804 friend muri |
| 11 | 105.5 | 23.1 | 676 | 1 ENV MLVFP | P26803 friend muri |
| 12 | 105 | 23.0 | 636 | 1 ENV MCFP | P15073 mink cell f |
| 13 | 105 | 23.0 | 661 | 1 ENV MLVCB | P08360 cas-br-e mu |
| 14 | 105 | 23.0 | 668 | 1 ENV FLVC6 | P21443 feline leuk |
| 15 | 104.5 | 22.9 | 675 | 1 ENV MLVF5 | P03390 friend muri |
| 16 | 103 | 22.5 | 211 | 1 ENV MSVFB | P03387 fbj murine |
| 17 | 103 | 22.5 | 586 | 1 ENV MPWF | P07575 simian maso |
| 18 | 103 | 22.5 | 587 | 1 ENV SVL1 | P03385 moloney mur |
| 19 | 103 | 22.5 | 665 | 1 ENV MLVMO | P03386 akv murine |
| 20 | 103 | 22.5 | 669 | 1 ENV MLVAV | P04502 kirsten mur |
| 21 | 102 | 22.3 | 201 | 1 ENV MLVKI | P06752 feline leuk |
| 22 | 102 | 22.3 | 639 | 1 ENV FLVSA | P21436 homuly muri |
| 23 | 102 | 22.3 | 666 | 1 ENV MLVHO | P11268 radiation m |
| 24 | 101 | 22.1 | 665 | 1 ENV MLVRD | P31794 radiation m |
| 25 | 101 | 22.1 | 665 | 1 ENV MLVRK | P31796 avian splee |
| 26 | 97 | 21.2 | 567 | 1 ENV AVISN | P51520 simian retr |
| 27 | 95 | 20.8 | 574 | 1 ENV SRV2R | P21415 gibbon ape |
| 28 | 94 | 20.6 | 667 | 1 ENV GALV | P03388 mink cell f |
| 29 | 92 | 20.1 | 640 | 1 ENV MCFP3 | P03399 avian retic |
| 30 | 88 | 19.3 | 582 | 1 ENV AVIRE | P31791 feline endo |
| 31 | 84 | 18.4 | 671 | 1 ENV FENV1 | P21412 squirrel mo |
| 32 | 71 | 15.5 | 575 | 1 ENV SRVH | Q9ukf5 homo sapien |
| 33 | 67.5 | 14.8 | 820 | 1 AD29_HUMAN | |

RESULT 1

ALIGNMENTS

| | | | | | | |
|-----|------|------|------|---|-------------|--------------------|
| 34 | 66.5 | 14.6 | 594 | 1 | AAP1_SCHPO | Q92367 schizosacch |
| 35 | 66.5 | 14.6 | 1255 | 1 | PER2_HUMAN | O15055 homo sapien |
| 36 | 66 | 14.4 | 359 | 1 | KLP8_HUMAN | O95600 homo sapien |
| 37 | 66 | 14.4 | 448 | 1 | PRS4_ORYSA | P46466 oryza sativ |
| 38 | 65 | 14.2 | 676 | 1 | RLM1_YEAST | Q12224 saccharomyc |
| 39 | 65 | 14.2 | 1321 | 1 | PGCN_HUMAN | O14594 homo sapien |
| 40 | 64.5 | 14.1 | 337 | 1 | INGS_HUMAN | P38484 homo sapien |
| 41 | 64.5 | 14.1 | 637 | 1 | P73_GERAE | O9xak8 cercopithe |
| 42 | 64 | 14.0 | 193 | 1 | ACD1_XANAC | P58902 xanthomonas |
| 43 | 64 | 14.0 | 867 | 1 | VL96_IRV1 | P22856 tipula irid |
| 44 | 64 | 14.0 | 1206 | 1 | FM14_MOUSE | Q05859 mus musculu |
| 45 | 63.5 | 13.9 | 915 | 1 | PDB2_ATHAT | O23078 arabidopsis |
| 46 | 63 | 13.8 | 1257 | 1 | PER2_MOUSE | O54943 mus musculu |
| 47 | 63 | 13.8 | 1284 | 1 | NRCA_CHICK | P35331 gallus gall |
| 48 | 62.5 | 13.7 | 357 | 1 | CYCK_HUMAN | O75909 homo sapien |
| 49 | 62.5 | 13.7 | 636 | 1 | P73_HUMAN | O13350 homo sapien |
| 50 | 62.5 | 13.7 | 831 | 1 | NFH_RAT | P16884 rattus norv |
| 51 | 62 | 13.6 | 216 | 1 | CAM2_BRARE | Q90479 brachydanio |
| 52 | 62 | 13.6 | 470 | 1 | SSF1_MOUSE | Q91y8 mus musculu |
| 53 | 62 | 13.6 | 1277 | 1 | CAML_FUGRU | Q98902 fugu rubrip |
| 54 | 62 | 13.6 | 1304 | 1 | NRCA_HUMAN | Q92823 homo sapien |
| 55 | 62 | 13.6 | 2064 | 1 | SIF1_DROME | P91621 drosophila |
| 56 | 61.5 | 13.5 | 110 | 1 | THI2_ANASP | P20857 anabaena sp |
| 57 | 61.5 | 13.5 | 533 | 1 | RAG2_ONCMY | Q91193 oncornynch |
| 58 | 61.5 | 13.5 | 1103 | 1 | AKA2_HUMAN | Q9y2d5 homo sapien |
| 59 | 61 | 13.3 | 257 | 1 | MOAB_KLEAE | P54795 klebsiella |
| 60 | 61 | 13.3 | 406 | 1 | PG2_HUMAN | Q9b7g0 homo sapien |
| 61 | 61 | 13.3 | 447 | 1 | AMEN_HUMAN | Q9np70 homo sapien |
| 62 | 61 | 13.3 | 721 | 1 | ENF1_TORCA | P14400 torpedo cal |
| 63 | 61 | 13.3 | 1300 | 1 | SAL3_HUMAN | Q9bxa9 homo sapien |
| 64 | 60.5 | 13.2 | 357 | 1 | RUVB_STRCO | Q91291 streptomyce |
| 65 | 60.5 | 13.2 | 5376 | 1 | ZAN_MOUSE | Q88799 mus musculu |
| 66 | 60 | 13.1 | 252 | 1 | Y8B5_PSEAE | Q9hvx2 pseudomonas |
| 67 | 59.5 | 13.0 | 290 | 1 | LEP4_PSEAE | P22610 pseudomonas |
| 68 | 59.5 | 13.0 | 609 | 1 | VE1_HPV49 | P36729 human papil |
| 69 | 59.5 | 13.0 | 633 | 1 | XRC1_HUMAN | P18887 homo sapien |
| 70 | 59.5 | 13.0 | 875 | 1 | CNSA_HUMAN | O78074 homo sapien |
| 71 | 59.5 | 13.0 | 1150 | 1 | RGAI_SCHPO | Q4052 schizosacch |
| 72 | 59.5 | 13.0 | 2696 | 1 | NSD1_HUMAN | Q96173 homo sapien |
| 73 | 59 | 12.9 | 270 | 1 | CPCL_NEUCR | P11115 neurospora |
| 74 | 59 | 12.9 | 360 | 1 | RLPA_VERPE | Q8zdg6 versinia pe |
| 75 | 59 | 12.9 | 567 | 1 | SIX5_MOUSE | P70178 mus musculu |
| 76 | 59 | 12.9 | 797 | 1 | FTSK_STREP | O8cmms staphylococ |
| 77 | 59 | 12.9 | 861 | 1 | RFC1_YEAST | P38530 saccharomyc |
| 78 | 59 | 12.9 | 1044 | 1 | BUB1_SCHPO | O94751 schizosacch |
| 79 | 59 | 12.9 | 1054 | 1 | CYPE_BACSU | O08336 bacillus su |
| 80 | 58.5 | 12.8 | 149 | 1 | YE65_ARCFU | O28807 archaeoglob |
| 81 | 58.5 | 12.8 | 463 | 1 | SYE1_HELPY | Q9z1z7 helicobacte |
| 82 | 58.5 | 12.8 | 483 | 1 | YAE1_SCHPO | P96551 helicobacte |
| 83 | 58.5 | 12.8 | 777 | 1 | YAE1_SCHPO | Q09697 schizosacch |
| 84 | 58.5 | 12.8 | 1170 | 1 | YKD8_YEAST | P32862 saccharomyc |
| 85 | 58.5 | 12.8 | 1603 | 1 | PSC_DROME | P35820 drosophila |
| 86 | 58 | 12.7 | 273 | 1 | SCF_HUMAN | P21583 homo sapien |
| 87 | 58 | 12.7 | 274 | 1 | FRT1_MOUSE | P70339 mus musculu |
| 88 | 58 | 12.7 | 395 | 1 | EFTU_BACST | O50306 bacillus st |
| 89 | 58 | 12.7 | 610 | 1 | SPPA_SYNY3 | P73689 synchocyst |
| 90 | 58 | 12.7 | 647 | 1 | CPN4_HUMAN | Q96m77 homo sapien |
| 91 | 58 | 12.7 | 663 | 1 | TFB1_NEUCR | Q9psn7 neurospora |
| 92 | 58 | 12.7 | 699 | 1 | AT6B_MOUSE | O35451 mus musculu |
| 93 | 58 | 12.7 | 739 | 1 | IMMT_DROME | P91928 drosophila |
| 94 | 58 | 12.7 | 877 | 1 | SULT_HSCHPO | O74377 schizosacch |
| 95 | 58 | 12.7 | 1058 | 1 | POL_HV2D2 | P15833 human immun |
| 96 | 58 | 12.7 | 1289 | 1 | SIP3_YEAST | P38717 saccharomyc |
| 97 | 58 | 12.7 | 1790 | 1 | SPR1_DROME | O8mgw8 drosophila |
| 98 | 57.5 | 12.6 | 335 | 1 | OXAA_BIFLO | O89616 bifidobacte |
| 99 | 57.5 | 12.6 | 357 | 1 | Y264_LISIN | Q92f41 listeria in |
| 100 | 57.5 | 12.6 | 402 | 1 | STAC_HUMAN | Q99469 homo sapien |

ENV_FLVGL
ID ENV_FLVGL STANDARD; PRT; 642 AA.
AC P08359;
DT 01-AUG-1988 (Rel. 08, Last created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein
DE GP70; Spike protein P15E].
GN ENV.
OS Feline leukemia virus (strain A/Glasgow-1).
OS Viruses; Retroviridae; Gammaretrovirus.
OC NCBI_TaxID=11769;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=86200439; PubMed=3009890;
RX Stewart M.A., Warnock M., Wheeler A., Wilkie N., Mullins J.I.,
RA Onions D.E., Neil J.C.;
RT "Nucleotide sequences of a feline leukemia virus subgroup A envelope
RT gene and long terminal repeat and evidence for the recombinational
RT origin of subgroup B viruses.";
RL J. Virol. 58:825-834 (1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M12500; AAA43053.1; -;
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 445 KNOB PROTEIN GP70.
FT CHAIN 446 642 SPIKE PROTEIN P15E.
FT CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 642 AA; 71053 MW; 2DEFS99EFC245EC CRC64;
Query Match 24.1%; Score 110; DB 1; Length 642;
Best Local Similarity 43.1%; Pred. No. 0.00074;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;
QY 1 ILPLFGPLAAILLLFGFCIFNLVNFVSSRIEAVKLMQEKMSKTKIYRRPLDRP 58
DB 587 ISSINGPLLLILLFGFCILNRLVQFVKDRISVVQALILTQQYQKQY--DPDRP 642
RESULT 2
ENV_FSVGA
ID ENV_FSVGA STANDARD; PRT; 662 AA.
AC P03391; F21446;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein
DE GP70; Spike protein P15E; R protein].
GN ENV.
OS Feline sarcoma virus (strain Gardner-Arnstein) (Ga-FeSV) (Gardner-
OS Arnstein feline leukemia oncovirus B).
OC NCBI_TaxID=11769;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=86200439; PubMed=3009890;
RX Stewart M.A., Warnock M., Wheeler A., Wilkie N., Mullins J.I.,
RA Onions D.E., Neil J.C.;
RT "Nucleotide sequences of a feline leukemia virus subgroup A envelope
RT gene and long terminal repeat and evidence for the recombinational
RT origin of subgroup B viruses.";
RL J. Virol. 58:825-834 (1986).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M12500; AAA43053.1; -;
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 445 KNOB PROTEIN GP70.
FT CHAIN 446 642 SPIKE PROTEIN P15E.
FT CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 334 334 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 642 AA; 71053 MW; 2DEFS99EFC245EC CRC64;

Viruses; Retroviridae; Gammaretrovirus.
OC NCBI_TaxID=11774;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=84115095; PubMed=6319767;
RX Nunberg J.H., Williams M.E., Innis M.A.;
RA "Nucleotide sequences of the envelope genes of two isolates of feline
RT leukemia virus subgroup B.";
RL J. Virol. 49:629-632 (1984).
CC -----
CC SEQUENCE FROM N.A.
RP MEDLINE=83216322; PubMed=6304347;
RX Elder J.H., Mullins J.I.;
RA "Nucleotide sequence of the envelope gene of Gardner-Arnstein feline
RT leukemia virus B reveals unique sequence homologies with a murine
RT mink cell focus-forming virus.";
RL J. Virol. 46:871-880 (1983).
CC -----
CC SEQUENCE FROM N.A.
RP MEDLINE=84131936; PubMed=6321156;
RX Wunsch M., Schulz A.S., Koch W., Friedrich R., Hunsmann G.;
RA "Sequence analysis of Gardner-Arnstein feline leukaemia virus
RT envelope gene reveals common structural properties of mammalian
RT retroviral envelope genes.";
RL EMBO J. 2:2239-2246 (1983).
CC -----
CC SEQUENCE FROM N.A.
RP MEDLINE=88044502; PubMed=2823466;
RX Guilhot S., Hampe A., D'Auriol L., Galibert F.;
RA "Nucleotide sequence analysis of the LTRs and env genes of SM-FeSV
RT and GA-FeSV.";
RL Virology 161:252-258 (1987).
CC -----
CC -!- P1M: Specific enzymatic cleavages in vivo yield mature proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K01209; AAA43052.1; -;
DR EMBL; V01172; CAA24497.1; -;
DR EMBL; X00188; CAA25008.1; -;
DR EMBL; M03026; -; NOT_ANNOTATED_CDS.
DR PIR; A03991; VCVWGF.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 465 KNOB PROTEIN GP70.
FT CHAIN 466 645 SPIKE PROTEIN P15E.
FT CHAIN 646 662 R PROTEIN.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 15 15 MISSING (IN REF. 4).
FT CONFLICT 15 15 V -> I (IN REF. 4).
FT CONFLICT 41 41 T -> V (IN REF. 4).
FT CONFLICT 47 47 LVGTGK -> VQVNTQ (IN REF. 4).
FT CONFLICT 51 56 FPMIF -> YPIILHV (IN REF. 4).
FT CONFLICT 70 75 IIGTNWPSQDEPPFG -> LVGDSWEPIVDPNNVKGAR
FT CONFLICT 80 95 YSSSK (IN REF. 4).

```

D -> T (IN REF. 2).
LT -> SL (IN REF. 2).
T -> R (IN REF. 2).
D -> P (IN REF. 2).
LPPTTPPOAGTG-> HSHQTSTGGDR (IN REF. 2).
R -> K (IN REF. 2). CRC64;
SEQUENCE 679 AA; 74453 MW; 5242BD0DD86554039 CRC64;

Query Match      23.9%; Score 109; DB 1; Length 679;
Best Local Similarity 32.5%; Pred. No. 0.001;
Matches 26; Conservative 12; Mismatches 22; Indels 20; Gaps 2;

          1 ILPFLGLAAIILLLFGPCINLLWNFVSSRIEAVKLQMEPKWSKTKIVRRPLDRPAS 60
            :|||:|||:|||||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
        620 ISTMGPLVILLILLLFGPCINRLNVQFYKRISVVQALI-----LFOQY----- 664

          61 PRSDVNIDIKGTPEETISAQA 80
            |||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
        665 -----HQLKSIDPEEVESRE 679

RESULT 4
ENV BAEVM STANDARD; PRT; 563 AA.
ID ENV BAEVM STANDARD; PRT; 563 AA.
AC P10269; 1989 (Rel. 10, Created)
01-MAR-1989 (Rel. 10, Last annotation update)
11-MAR-1989 (Rel. 10, Last annotation update)
16-OCT-2001 (Rel. 40, Last annotation update)
ENV polypeptide precursor (Coat polyprotein) [Contains: Outer membrane protein GP70; Transmembrane protein P20E].
DE DE
DE ENV
GN GN
OS Baboon endogenous virus (strain M7).
OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
OX NCBI_TaxID=11764;
RN RN
[1]
SEQ FROM N.A.K.
Kato S., Matsuo K., Nishimura N., Takahashi N., Takano T.;
"The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retroviruses."; Jpn. J. Genet. 62:127-137(1987).
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions in its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; D10032; BAA0924.1; --
EMBL; X05470; CAA29028.1; --
PIR; JT0262; VCMVM7.
HSP; F03385; IMOF.
InterPro; IPR002050; Env_polyprotein.
Pfam; PF00429; Env_polyprotein_1.
Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 20
FT CHAIN 21 346 OUTER MEMBRANE PROTEIN GP70 (POTENTIAL).
FT FT 347 563 TRANSMEMBRANE PROTEIN P20E (POTENTIAL).
FT CHAIN 113 113 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 229 229 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 264 264 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 282 282 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 292 292 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 306 306 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 312 312 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 321 321 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 339 339 N-LINKED (GLCNAC...) (POTENTIAL).
FT FT CARBOHYD 469 469 N-LINKED (GLCNAC...) (POTENTIAL).
SEQUENCE 563 AA; 61879 MW; 9573137DC4620BE7 CRC64;

```


or send an email to license@isb-sib.ch.

EMBL; V01201; CAA24517.1; --
 PIR; A03982; A03982; Env_polyprotein.
 InterPro; IPR002050; Env_polyprotein; 1.
 Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein.
 FT NON TER
 SQ SEQUENCE 77 AA; 8984 MW; 91A54FCE9B659BC CRC64;
 Query Match 23.1%; Score 105.5; DB 1; Length 77;
 Best Local Similarity 50.0%; Pred. No. 0.0002;
 Matches 21; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

QY 6 GPLAAILLLFGPCIFNLLVNFVSSRIEAVK-LQMEPKQMS 46
 DB 29 GPLLLLLLLILGPCIINRLVQFINRVSAVKILVLRQYQT 70

RESULT 9
 ENV_RMCVF STANDARD; PRT; 640 AA.
 ID ENV_RMCVF STANDARD; PRT; 640 AA.
 AC P06745; Q85628; Q85629; Q89529;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polyprotein precursor (Coat polyprotein) [Contains: Knob protein
 GN ENV.
 OS Rauscher mink cell focus-inducing virus.
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11784;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85237696; PubMed=4009793;
 RA Vogt M., Hagblom C., Swift S., Haas M.;
 RT "Envelope gene and long terminal repeat determine the different
 RT biological properties of Rauscher, Friend, and Moloney mink cell
 RT focus-inducing viruses.";
 RL J. Virol. 55:184-192(1985).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; M10100; AAA46528.1; --
 PIR; A03990; VCMVRV.
 HSP; P03385; IMOF.
 InterPro; IPR002050; Env_polyprotein.
 InterPro; IPR008981; F_MuLV bind.
 Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 32
 FT CHAIN 33 443 KNOB PROTEIN GP70.
 FT CHAIN 444 623 COAT PROTEIN P12E.
 FT CHAIN 444 640 COAT PROTEIN P15E.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 339 339 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 640 AA; 70071 MW; 1E4450343643D799 CRC64;
 Query Match 23.1%; Score 105.5; DB 1; Length 640;
 Best Local Similarity 41.1%; Pred. No. 0.0022;
 Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;

Nicolaissen-Strouss K., Kumar H.P.M., Fitting T., Grant C.K.,
 Elder J.H.;
 "Natural feline leukemia virus variant escapes neutralization by a
 monoclonal antibody via an amino acid change outside the antibody-
 binding epitope.";
 J. Virol. 61:3410-3415(1987).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; J03448; AAA43048.1; --
 PIR; A27172; VCMVLB.
 HSP; P03385; IMOF.
 InterPro; IPR002050; Env_polyprotein.
 InterPro; IPR008981; F_MuLV bind.
 Pfam; PF00429; Env_polyprotein; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 465 KNOB PROTEIN GP70.
 FT CHAIN 466 662 SPIKE PROTEIN P15E.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT NON TER 662
 SQ SEQUENCE 662 AA; 73132 MW; D4AC18E197230575 CRC64;
 Query Match 23.4%; Score 107; DB 1; Length 662;
 Best Local Similarity 43.1%; Pred. No. 0.0016;
 Matches 22; Conservative 8; Mismatches 21; Indels 0; Gaps 0;

QY 1 ILPLGLAAILLLFGPCIFNLLVNFVSSRIEAVK-LQMEPKQMSKTKY 51
 DB 607 ISSIMGPLLLILLFGPCIINRLVQFINRVSAVKILVLRQYQT 657

RESULT 8
 ENV_SMSAV STANDARD; PRT; 77 AA.
 ID ENV_SMSAV STANDARD; PRT; 77 AA.
 AC P03384;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-AUG-1988 (Rel. 08, Last annotation update)
 DE P15E protein (Fragment).
 GN ENV.
 OS Simian sarcoma virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11817;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83144004; PubMed=6298772;
 RA Devare S.G., Reddy E.P., Law J.D., Robbins K.C., Aaronson S.A.;
 RT "Nucleotide sequence of the simian sarcoma virus genome:
 RT demonstration that its acquired cellular sequences encode the
 RT transforming gene product p28sis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:731-735(1983).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

QY 1 ILPLGLPAAIILLIFGFCIFNLLVNFVSSRIEAVK-----LQMEPK 43
DB 585 ISTINGPLIILLIFGFCILNRLVQVVKORISVVQALVLTQQYHQLKPLEYEPQ 640

RESULT 10
ENV_MLVFF STANDARD; PRT; 676 AA.
ID ENV_MLVFF
AC P26804;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor [Contains: Knob protein GP70; Spike protein P15E; R protein].
GN ENV.
OS Friend murine leukemia virus (isolate FB29) (F-MuLV).
OC Viruses; Retroviridae; Retroviridae; Mammalian type C retroviruses.
OX NCBI_TaxID=11797;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92107687; PubMed=1762923;
RA Perryman S., Nishio J., Chesbro B.;
RT "Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.";
RL Nucleic Acids Res. 19:6950-6950(1991).
CC -1- FUNCTION: INVOLVED IN HOST RANGE; VIRUS RECEPTOR BINDING; CELL FUSION.
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL; Z11128; CAA77479.1; --
DR PIR; S70395; S70395.
DR HSP; P03390; IAO.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 34
FT CHAIN 35 479 KNOB PROTEIN GP70.
FT CHAIN 480 659 SPIKE PROTEIN P15E.
FT CHAIN 660 676 R PROTEIN.
FT TRANSMEM 620 636 POTENTIAL.
FT DISULFID 80 132 BY SIMILARITY.
FT DISULFID 106 121 BY SIMILARITY.
FT DISULFID 107 117 BY SIMILARITY.
FT DISULFID 155 175 BY SIMILARITY.
FT DISULFID 167 180 BY SIMILARITY.
FT DISULFID 167 180 BY SIMILARITY.
FT DISULFID 212 218 BY SIMILARITY.
FT DISULFID 346 349 BY SIMILARITY.
FT DISULFID 395 407 BY SIMILARITY.
FT DISULFID 376 430 BY SIMILARITY.
FT DISULFID 437 450 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 676 AA; 74049 MW; E18C44511145188B CRC64;

Query Match 23.1%; Score 105.5; DB 1; Length 676;
Best Local Similarity 41.1%; Pred No. 0.0023;
Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;

```
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 676 AA; 18DB8A7FE0A8FA4 CRC64;

Query Match
Best Local Similarity 23.1%; Score 105.5; DB 1; Length 676;
Matches 23; Conservative 7; Mismatches 13; Indels 13; Gaps 1;

Qy 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 43
Db 621 ISTMGPLIILLFGPCILNRLVQFVKDRISVQALVLTQYHQLEPEYEQ 676

RESULT 12
ENV_MCFF
ID ENV_MCFF STANDARD; PRT; 636 AA.
AC P15073;
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
DE GP70; Spike protein P15E; R protein].
GN ENV.
OS Mink cell focus-forming murine leukemia virus.
OC Viruses; Retroviridae; Retroviridae.
OX NCBI_TaxID=11935;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89085614; PubMed=2535909;
RA Chattopadhyay S.K., Baroudy B.M., Holmes K.L., Fredrickson T.N.,
RA Lander M.R., Morse H.C. III, Hartley J.W.;
RT "Biologic and molecular genetic characteristics of a unique MCF virus
RT that is highly leukemogenic in ecotropic virus-negative mice.";
RL Virology 168:90-100(1989).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M23029; -; NOT_ANNOTATED_CDS.
CC HSP; P03385; IMOF.
CC InterPro; IPR002050; Env polypeptide.
CC InterPro; IPR008981; F_MuLV_bind.
CC Pfam; PF00429; ENV_polyprotein; 1.
CC Coat protein; Glycoprotein; Polyprotein; Signal.
CC SIGNAL 1 30
CC CHAIN 31 440 KNOB PROTEIN GP70.
CC CHAIN 441 620 SPIKE PROTEIN P15E.
CC CHAIN 621 636 R PROTEIN.
CC FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 636 AA; 69113 MW; DF6208F7EA968E2A CRC64;

Query Match
Best Local Similarity 23.0%; Score 105; DB 1; Length 636;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 582 ISTMGPLIILLFGPCILNRLVQFVKDRISVQ 618

RESULT 13
```

```
ENV_MLVCB
ID ENV_MLVCB STANDARD; PRT; 661 AA.
AC P08360;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
DE GP70; Spike protein P15E; R protein].
GN ENV.
OS Cas-Br-E murine leukemia virus.
OC Viruses; Retroviridae; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11792;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87061215; PubMed=3023680;
RA Rassart E., Nelbach L., Jolicoeur P.;
RT "Cas-Br-E murine leukemia virus: sequencing of the paralytogenic
RT region of its genome and derivation of specific probes to study its
RT origin and the structure of its recombinant genomes in leukemic
RT tissues.";
RL J. Virol. 60:910-919(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91227170; PubMed=1840655;
RA Perryman S.M., McAtee F.J., Portis J.L.;
RT "Complete nucleotide sequence of the neurotropic murine retrovirus
RT CAS-BR-E.";
RL Nucleic Acids Res. 19:1707-1707(1991).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M14702; AAA46512.1; -
CC EMBL; X57540; -; NOT_ANNOTATED_CDS.
CC PIR; B26103; VCMVCB.
CC HSP; P03385; IMOF.
CC InterPro; IPR002050; Env polypeptide.
CC InterPro; IPR008981; F_MuLV_bind.
CC Pfam; PF00429; ENV_polyprotein; 1.
CC Coat protein; Glycoprotein; Polyprotein; Signal.
CC SIGNAL 1 465
CC CHAIN 466 645 KNOB PROTEIN GP70.
CC CHAIN 646 661 SPIKE PROTEIN P15E.
CC FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 322 322 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 476 477 PE -> LG (IN REF. 2).
CC SEQUENCE 661 AA; 72624 MW; 255BEE3AF62FB9FF CRC64;

Query Match
Best Local Similarity 23.0%; Score 105; DB 1; Length 661;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 607 ISTMGPLIILLFGPCILNRLVQFVKDRISVQ 643

RESULT 14
ENV_FLVC6
ID ENV_FLVC6 STANDARD; PRT; 668 AA.
AC P21443;
DT 01-MAY-1991 (Rel. 18, Created)
```


01-MAY-1991 (Rel. 18, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 ENV polypeptide precursor [Contains: Coat protein GP70; Coat protein P15E].
 ENV.
 OS Feline leukemia provirus (clone CFE-6).
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11922;
 RN [1]
 PP SEQUENCE FROM N.A.
 RX MEDLINE=89199802; PubMed=2539525;
 RA Kumar D.V., Berry B.T., Roy-Burman P.;
 RT "Nucleotide sequence and distinctive characteristics of the env gene
 of endogenous feline leukemia provirus.";
 RL J. Virol. 63:2379-2384(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M25425; AAA30809.1; --
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env polypeptide.
 DR InterPro; IPR008981; F MuLV bind.
 DR Pfam; PF00429; ENV polypeptide; 1.
 KW Coat protein; Glycoprotein; Polypeptide; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 465 COAT PROTEIN GP70.
 FT CHAIN 466 668 COAT PROTEIN P15E.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 668 AA; 74298 MW; 4A6A06CF2EB8CE25 CRC64;
 Query Match 23.0%; Score 105; DB 1; Length 668;
 Best Local Similarity 39.4%; Pred. No. 0.0026;
 Matches 26; Conservative 11; Mismatches 25; Indels 4; Gaps 2;
 Qy 1 ILPLGLPAAIILLLLFGPCIFNLLVSSRIEAVKLQWEPKQSKTIYR-RPLDREA 59
 Db 606 ISSMGLPMLILLLLFGPCILNRLVQFVKDRISVVQTLVLTQYQRLGQWLRP---TV 662
 Qy 60 SPRSDV 65
 Db 663 SPQLNV 668
 RESULT 15
 ENV_MLVP5
 ID ENV_MLVP5 STANDARD; PRT; 675 AA.
 AC P03390;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ENV polypeptide precursor [Contains: Knob protein GP70; Spike protein
 P15E; R protein].
 GN ENV.
 OS Friend murine leukemia virus (isolate 57) (F-MuLV).
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11796;
 RN [1]
 PP SEQUENCE FROM N.A.

BA Friedrich R.W., Koch W., von Maydell-Livonius U., Schrewe H.,
 RA Zimmermann W.;
 RL Submitted (XX-1991) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=84138778; PubMed=6321768;
 RA Koch W., Zimmermann W., Oliff A., Friedrich R.W.;
 RT "Molecular analysis of the envelope gene and long terminal repeat of
 Friend mink cell focus-inducing virus: implications for the functions
 of these sequences";
 RL J. Virol. 49:828-840(1984).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 43-270.
 RX MEDLINE=97433384; PubMed=9287219;
 RA Pass D., Davey R.A., Hanson C.A., Kim P.S., Cunningham J.M.,
 RA Berger J.M.;
 RT "Structure of a murine leukemia virus receptor-binding glycoprotein
 at 2.0-A resolution";
 RL Science 277:1662-1666(1997).
 CC -!- FUNCTION: INVOLVED IN HOST RANGE; VIRUS RECEPTOR BINDING; CELL
 FUSION.
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X02794; CAA26561.1; --
 DR PDB; IACI; 1S-OCT-97.
 DR InterPro; IPR002050; Env polypeptide.
 DR InterPro; IPR008981; F MuLV bind.
 DR Pfam; PF00429; ENV polypeptide; 1.
 KW Coat protein; Glycoprotein; Polypeptide; Transmembrane; Signal;
 3D-structure; Zinc; Metal-binding.
 FT SIGNAL 1 34 BY SIMILARITY.
 FT CHAIN 35 479 KNOB PROTEIN GP70.
 FT CHAIN 480 659 SPIKE PROTEIN P15E.
 FT CHAIN 660 675 R PROTEIN.
 FT TRANSMEM 620 636 POTENTIAL.
 FT METAL 55 55 ZINC.
 FT DISULFID 80 132
 FT DISULFID 106 121
 FT DISULFID 107 117
 FT DISULFID 155 175
 FT DISULFID 167 180
 FT DISULFID 212 218
 FT DISULFID 346 349
 FT DISULFID 395 407
 FT DISULFID 376 430
 FT DISULFID 437 450
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 336 336 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT STRAND 44 52
 FT TURN 54 55
 FT STRAND 58 67
 FT TURN 68 69
 FT STRAND 75 75
 FT STRAND 78 78
 FT STRAND 79 83
 FT HELIX 84 85
 FT TURN 88 90
 FT HELIX 93 94
 FT TURN 97 98
 FT STRAND 106 106

FT TURN 107
 FT TURN 110
 FT TURN 111
 FT TURN 116
 FT STRAND 118
 FT HELIX 119
 FT TURN 122
 FT TURN 126
 FT TURN 135
 FT TURN 138
 FT TURN 148
 FT TURN 152
 FT TURN 158
 FT TURN 162
 FT HELIX 164
 FT HELIX 167
 FT TURN 170
 FT TURN 173
 FT TURN 178
 FT TURN 183
 FT TURN 187
 FT TURN 188
 FT STRAND 197
 FT HELIX 201
 FT TURN 205
 FT TURN 213
 FT TURN 214
 FT STRAND 220
 FT HELIX 227
 FT TURN 234
 FT TURN 237
 FT STRAND 239
 FT TURN 245
 FT TURN 248
 FT STRAND 253
 FT HELIX 266
 SQ SEQUENCE 675 AA; A0970385422BB3D3 CRC64;
 Query Match 22.9%; Score 104.5; DB 1; Length 675;
 Best Local Similarity 41.8%; Pred. No. 0.0029;
 Matches 23; Conservative 6; Mismatches 13; Indels 13; Gaps 1;
 QY 1 ILPLGLPAAIILLLLFGPCIFNLVNFVSRIRAVK-----LQMEP 42
 DB 621 ISTINGPLIILLLLFGPCILNRLVQFKDRISVWQALVLTQOYHOLKPLEYEP 675
 RESULT 16
 ENV MSVFB STANDARD; PRT; 211 AA.
 AC P03387;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide (Coat polypeptide) [Contains: Knob protein GP70; Spike protein P15E; R protein] (Fragment).
 DE protein P15E; R protein] (Fragment).
 GN ENV.
 OS FBJ murine osteosarcoma virus.
 OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=11805;
 RN [1]
 RP SEQUENCE FROM N.A. MEDLINE=83180421; PubMed=6301687;
 RX "Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini.";
 RL Cell 32:1241-1255(1983).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; M12349; AAA47712.1; -.
 DR PIR; D25839; VCLJWP.
 DR HSSP; P03385; 1MOP.
 DR InterPro; IPR002050; Env_polypeptide.
 DR Pfam; PF00429; Env_polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 394
 FT CHAIN 395 586
 FT CARBOHYD 120 130
 FT CARBOHYD 237 237
 FT CARBOHYD 264 264
 FT CARBOHYD 276 276
 FT CARBOHYD 291 291
 FT CARBOHYD 304 304
 FT CARBOHYD 318 318
 FT CARBOHYD 324 324
 FT CARBOHYD 339 339
 FT CARBOHYD 357 357
 FT CARBOHYD 487 487
 SQ SEQUENCE 586 AA; 63882 MW; 3CE7A399D9E2F450 CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 586;
 Best Local Similarity 45.9%; Pred. No. 0.0036;
 Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;
 DR EMBL; V01184; CAA24506.1; -.
 DR PIR; A03986; A03986.
 DR HSSP; P03385; 1MOP.

DR InterPro; IPR002050; Env_polypeptide.
 DR Pfam; PF00429; Env_polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein.
 FT NON_TER 1 1
 FT CHAIN <1 10
 FT CHAIN 11 190
 FT CHAIN 191 211
 FT CHAIN 191 211
 SQ SEQUENCE 211 AA; 23708 MW; 082EC960ED0E185B CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 211;
 Best Local Similarity 51.4%; Pred. No. 0.0011;
 Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
 QY 1 ILPLGLPAAIILLLLFGPCIFNLVNFVSRIRAVK 37
 DB 152 ISTINGPLIILLLLFGPCILNRLVQFKDRISVWQ 188
 RESULT 17
 ENV MPV STANDARD; PRT; 586 AA.
 AC P07575;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein GP70; Coat protein GP20].
 DE GP70; Coat protein GP20].
 GN ENV.
 OS Simian Mason-Pfizer virus (MPV).
 OC Viruses; Retroviral viruses; Retroviridae; Betaretrovirus.
 OX NCBI_TaxID=11855;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE 6A).
 RX MEDLINE=86189951; PubMed=2421920;
 RA Sonigo P., Barker C., Hunter E., Wain-Hobson S.;
 RT "Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus.";
 RL Cell 45:375-385(1986).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL; M12349; AAA47712.1; -.
 DR PIR; D25839; VCLJWP.
 DR HSSP; P03385; 1MOP.
 DR InterPro; IPR002050; Env_polypeptide.
 DR Pfam; PF00429; Env_polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 394
 FT CHAIN 395 586
 FT CARBOHYD 120 130
 FT CARBOHYD 237 237
 FT CARBOHYD 264 264
 FT CARBOHYD 276 276
 FT CARBOHYD 291 291
 FT CARBOHYD 304 304
 FT CARBOHYD 318 318
 FT CARBOHYD 324 324
 FT CARBOHYD 339 339
 FT CARBOHYD 357 357
 FT CARBOHYD 487 487
 SQ SEQUENCE 586 AA; 63882 MW; 3CE7A399D9E2F450 CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 586;
 Best Local Similarity 45.9%; Pred. No. 0.0036;
 Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPFLGLAAIILLLLFGFCIFNLLVNFVSSRIEAVK 37
 Db 526 VMLPLGLCLLLVLSFGPIIFNKLMTFKHQIESIQ 562

RESULT 18

ENV_SRV1 STANDARD; PRT; 587 AA.
 AC F04027;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Coat protein
 DE GP70; Coat protein GP20].
 GN ENV.
 OS Simian retrovirus SRV-1.
 OC Viruses; Retroviral viruses; Retroviridae; Betaretrovirus.
 OC NCBI_TaxID=11942;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8615168; PubMed=3006247;
 RA Power M.D., Marx P.A., Bryant M.L., Gardner M.B., Barr P.J.,
 RA Luciw P.A.;
 RT "Nucleotide sequence of SRV-1, a type D simian acquired immune
 RT deficiency syndrome retrovirus.";
 RL Science 231:1567-1572(1986).
 CC -!- P.TM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL; M1841; AAA47733.1; -.
 DR HSP; P03385; IMOF.
 DR InterPro; IPR002050; Env. polypeptide.
 DR Pfam; PF00429; Env. polypeptide; 1.
 KW Coat protein; Glycoprotein; Polypeptide; Transmembrane; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 395 COAT PROTEIN GP70.
 FT CHAIN 396 587 COAT PROTEIN GP20.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 528 548 POTENTIAL.
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 587 AA; 64474 MW; 80939DD3BFB65A8D CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 587;
 Best Local Similarity 45.8%; Pred. No. 0.0036;
 Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPFLGLAAIILLLLFGFCIFNLLVNFVSSRIEAVK 37
 Db 527 VMLPLGLCLLLVLSFGPIIFNKLMTFKHQIESIQ 563

RESULT 19

ENV_MLVMO STANDARD; PRT; 665 AA.
 AC P03385;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein
 DE GP70; Spike protein P15E; R protein].
 GN ENV.
 OS Moloney murine leukemia virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OC NCBI_TaxID=11801;
 RN [1]
 RP SEQUENCE FROM N.A. (CLONE PMLV-1).
 RX MEDLINE=82035843; PubMed=6169994;
 RA Shinnick T.M., Lerner R.A., Sutcliffe J.G.;
 RA "Nucleotide sequence of Moloney murine leukaemia virus.";
 RL Nature 293:543-548(1981).
 RN [2]
 RP SEQUENCE OF 496-665 FROM N.A. (PROVIRUS).
 RX MEDLINE=81052384; PubMed=6159543;
 RA Sutcliffe J.G., Shinnick T.M., Green N., Liu F.-T., Niman H.L.,
 RA Lerner R.A.;
 RA "Chemical synthesis of a polypeptide predicted from nucleotide
 RA sequence allows detection of a new retroviral gene product.";
 RL Nature 287:801-805(1980).
 RN [3]
 RP SEQUENCE OF 484-665 FROM N.A. (PROVIRUS; CLONE PMLV-201).
 RX MEDLINE=81013872; PubMed=6251454;
 RA Sutcliffe J.G., Shinnick T.M., Verma I.M., Lerner R.A.;
 RA "Nucleotide sequence of Moloney leukemia virus: 3' end reveals
 RA details of replications, analogy to bacterial transposons, and an
 RA unexpected gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).
 RN [4]
 RP SEQUENCE OF 470-489 AND 598-665.
 RX MEDLINE=82082389; PubMed=6947213;
 RA Green N., Shinnick T.M., Witte O., Ponticelli A., Sutcliffe J.G.,
 RA Lerner R.A.;
 RA "Sequence-specific antibodies show that maturation of Moloney
 RA leukemia virus envelope polypeptide involves removal of a
 RA COOH-terminal peptide.";
 RL Proc. Natl. Acad. Sci. U.S.A. 78:6023-6027(1981).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 514-567.
 RX MEDLINE=96196884; PubMed=8612076;
 RA Fass D., Harrison S.C., Kim P.S.;
 RA "Retrovirus envelope domain at 1.7-A resolution.";
 RL Nat. Struct. Biol. 3:465-469(1996).
 CC -!- P.TM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; J02255; AAB59943.1; -.
 CC PIR; A93265; VCVWEM.
 DR InterPro; IPR002050; Env. polypeptide.
 DR InterPro; IPR008981; F. MuLV bind.
 DR Pfam; PF00429; Env. polypeptide; 1.
 KW Glycoprotein; Coat protein; Polypeptide; Transmembrane; Signal;
 KW 3D-structure.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 34 469 KNOB PROTEIN GP70.
 FT CHAIN 470 649 SPIKE PROTEIN P15E.
 FT CHAIN 650 665 R PROTEIN.
 FT TRANSMEM 610 627 POTENTIAL.
 FT DISULFID 555 562

FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 655 655 Y -> F (IN REF. 3).
 FT CONFLICT 663 663 Y -> C (IN REF. 3 AND 4).
 FT HELIX 516 547
 FT TURN 548 548
 FT HELIX 549 551
 FT TURN 552 552
 FT HELIX 554 558
 FT TURN 559 559
 SQ SEQUENCE 665 AA; 73301 MW; 12EBA09C9FB93FE2 CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 665;
 Best Local Similarity 51.4%; Pred. No. 0.0041;
 Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
 QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
 DB 611 ISTMGPLVLLMILLFGPCILNRLVQFVKDRISVVQ 647
 RESULT 20
 ENV_MLVAV STANDARD; PRT; 669 AA.
 AC P03386;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide precursor (Coat polypeptide) [Contains: Knob protein GP70; Spike protein P15E; R protein].
 GN ENV.
 OS AKV murine leukemia virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11791;
 RN [1]_TaxID=11791;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84115072; PubMed=6319746;
 RA Herr W.;
 RT "Nucleotide sequence of AKV murine leukemia virus.";
 RL J. Virol. 49:471-478 (1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83090450; PubMed=6294621;
 RA Herr W., Corbin V., Gilbert W.;
 RT "Nucleotide sequence of the 3' half of AKV.";
 RL Nucleic Acids Res. 10:6931-6944 (1982).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82217034; PubMed=6283170;
 RA Lenz J., Crowther R., Straceski A., Haseltine W.;
 RT "Nucleotide sequence of the AKV env gene.";
 RL J. Virol. 42:519-529 (1982).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J01998; AA03092.1; -;
 CC EMBL; V01164; CA24493.1; ALT_INIT.
 CC PIR; A92995; VCVMEK.
 CC HSP; P03385; IMOF.
 CC InterPro; IPR002050; Env polypeptide.
 CC InterPro; IPR008981; F_MuLV_bind.

DR Pfam; PF00429; ENV polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein; Transmembrane; Signal.
 FT SIGNAL 1 31
 FT CHAIN 32 470 KNOB PROTEIN GP70.
 FT CHAIN 471 650 SPIKE PROTEIN P15E.
 FT CHAIN 651 669 R PROTEIN.
 FT TRANSMEM 473 493
 FT TRANSMEM 612 632 POTENTIAL.
 FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 35 35 E -> R (IN REF. 2).
 FT CONFLICT 463 463 E -> K (IN REF. 2).
 FT CONFLICT 592 592 E -> K (IN REF. 2).
 SQ SEQUENCE 669 AA; 73756 MW; 3A6C3845208A13F2 CRC64;
 Query Match 22.5%; Score 103; DB 1; Length 669;
 Best Local Similarity 51.4%; Pred. No. 0.0042;
 Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;
 QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
 DB 612 ISTMGPLVLLMILLFGPCILNRLVQFVKDRISVVQ 648
 RESULT 21
 ENV_MLVAV STANDARD; PRT; 201 AA.
 AC P04502;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ENV polypeptide (Coat polypeptide) [Contains: Knob protein GP70; Spike protein P15E; R protein] (Fragment).
 GN ENV.
 OS Kirsten murine leukemia virus.
 OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
 OX NCBI_TaxID=11800;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85014136; PubMed=6091040;
 RA Norton J.D., Connor J., Avery R.J.;
 RT "Genesis of Kirsten murine sarcoma virus: sequence analysis reveals recombination points and potential leukaemogenic determinant on parental leukaemia virus genome.";
 RL Nucleic Acids Res. 12:6839-6852 (1984).
 CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X00982; CAA25490.1; -;
 CC PIR; A03985; A03985.
 CC HSP; P03385; IMOF.
 CC InterPro; IPR002050; Env polypeptide.
 CC Pfam; PF00429; ENV polypeptide; 1.
 KW Coat protein; Glycoprotein; Polyprotein.
 FT CHAIN 1 1
 FT CHAIN <1 7 KNOB PROTEIN GP70.
 FT CHAIN 8 187 SPIKE PROTEIN P15E.
 FT CHAIN 188 201 R PROTEIN.
 SQ SEQUENCE 201 AA; 22389 MW; 3A900F7E3A0A3B15 CRC64;
 Query Match 22.3%; Score 102; DB 1; Length 201;

```

Best Local Similarity 54.5%; Pred. No. 0.0014;
Matches 18; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 5 LGPLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 153 MGPLIILLILLFGPCILNRLVQFVKDRISVQ 185

RESULT 22
ENV_FLVSA STANDARD; PRT; 639 AA.
AC P06752;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DE ENV polypeptide precursor (Coat polyprotein) [Contains: Knob protein
DE GP70; Spike protein P15E].
GN ENV.
OS Feline leukemia virus (strain C/Sarma).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=103919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86308240; PubMed=3018287;
RA Riedel N., Hoover E.A., Gasper P.W., Nicolson M.O., Mullins J.I.;
RT "Molecular analysis and pathogenesis of the feline aplastic anemia
RT retrovirus, feline leukemia virus C-Sarma.";
RL J. Virol. 60:242-250(1986).
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M14331; AAA43049.1; -.
DR PIR; A29013; VCMVSA.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal.
FT SIGNAL 1 33
FT CHAIN 34 442 KNOB PROTEIN GP70.
FT CARBOHYD 443 639 SPIKE PROTEIN P15E.
FT CARBOHYD 43 35 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 639 AA; 71161 MW; A309B64AC8EC74E4 CRC64;

Query Match 22.3%; Score 102; DB 1; Length 639;
Best Local Similarity 41.4%; Pred. No. 0.005;
Matches 24; Conservative 8; Mismatches 24; Indels 2; Gaps 1;

QY 1 ILPLGLAAILLLFGPCIFNLLVNFVSSRIEAVKLMQEPKMSKTYRPLDRP 58
DB 584 ISSIMGPLIILLILLFGPCILNRLVQFVKDRISVQALI--LTQYQOQIQYDSRDP 639

RESULT 23
ENV_MLVHO STANDARD; PRT; 566 AA.
ID ENV_MLVHO
AC P21436;
DT 01-MAY-1991 (Rel. 18, Created)

```

```

DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polypeptide precursor [Contains: Knob protein GP76; Spike protein
DE P15E].
GN ENV.
OS HomULV murine leukemia virus (Mus hortulanus virus).
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11799;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90051094; PubMed=25545479;
RA Voytek P., Kozak C.A.;
RT "Nucleotide sequence and mode of transmission of the wild mouse
RT ecotropic virus, HomULV.";
RL Virology 173:58-67(1989).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M26527; -. NOT ANNOTATED_CDS.
DR PIR; B32594; VCMVHL.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Coat protein; Glycoprotein; Polyprotein; Signal; Transmembrane.
FT SIGNAL 1 30
FT CHAIN 31 466 KNOB PROTEIN GP76.
FT CHAIN 467 666 SPIKE PROTEIN P15E.
FT TRANSMEM 473 489 POTENTIAL.
FT TRANSMEM 607 623 POTENTIAL.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 666 AA; 73034 MW; FDC77956E4B213D1 CRC64;

Query Match 22.3%; Score 102; DB 1; Length 666;
Best Local Similarity 51.4%; Pred. No. 0.0053;
Matches 19; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 ILPLGLAAILLLFGPCIFNLLVNFVSSRIEAVK 37
DB 608 ISTIMGPLIILLILLFGPCILNRLVQFVKDRISVQ 644

RESULT 24
ENV_MLVSD STANDARD; PRT; 665 AA.
ID ENV_MLVSD
AC P11268;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE ENV polypeptide precursor (Coat polyprotein) [Contains: Knob protein
DE GP76; Spike protein P15E; R protein].
GN ENV.
OS Radiation murine leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11787;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87207680; PubMed=3033897;
RA Merregaert J., Janowski M., Reddy E.P.;
RT "Nucleotide sequence of a radiation leukemia virus genome.";
RL Virology 158:88-102(1987).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K03363; AAA46519.1; ALT_INIT.
DR PIR; C26183; VCMVVR.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Glycoprotein; Coat protein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 31
FT CHAIN 32 467 KNOB PROTEIN GP76.
FT CHAIN 468 644 SPIKE PROTEIN P15E.
FT CHAIN 645 665 R PROTEIN.
FT TRANSMEM 605 621 POTENTIAL.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 665 AA; 73083 MW; 93B2BFCC9E45984 CRC64;
Query Match 22.1%; Score 101; DB 1; Length 665;
Best Local Similarity 31.2%; Pred No. 0.0067;
Matches 25; Conservative 12; Mismatches 23; Indels 20; Gaps 2;
QY 1 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKYRRLDRPAS 60
Db 606 ISTINGPLVLLILLLLGPCIILNRLVQFVKDRISVVQALV-----LTQQY----- 650
QY 61 PRSDVNDIKGTPPEISAAQ 80
Db 651 -----HQLKSIDPEMESRE 665
RESULT 25
ENV_MLVK STANDARD; PRT; 665 AA.
AC F31794;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ENV polyprotein precursor (Coat polyprotein) [Contains: Coat protein
DE GP76; Coat protein P15E; R protein].
GN ENV.
OS Radiation murine leukemia virus (strain Kaplan).
OC Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=31689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9233703; PubMed=1629969;
RA Poliquin L., Bergeron D., Fortier J.L., Paquette Y., Bergeron R.,
RA Rassart E.;
RT "Determinants of thymotropism in Kaplan radiation leukemia virus and
RT nucleotide sequence of its envelope region."
RL J. Virol. 66:5141-5146(1992).
CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M93052; AAA46526.1; -.
DR PIR; B42743; VCMVKA.
DR HSP; P03385; IMOF.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Glycoprotein; Coat protein; Polyprotein; Transmembrane; Signal.
FT SIGNAL 1 31
FT CHAIN 32 467 KNOB PROTEIN GP76.
FT CHAIN 468 644 SPIKE PROTEIN P15E.
FT CHAIN 645 665 R PROTEIN.
FT TRANSMEM 605 621 POTENTIAL.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 665 AA; 73085 MW; FA15ABGB0C63FOAA CRC64;
Query Match 22.1%; Score 101; DB 1; Length 665;
Best Local Similarity 31.2%; Pred No. 0.0067;
Matches 25; Conservative 12; Mismatches 23; Indels 20; Gaps 2;
QY 1 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKYRRLDRPAS 60
Db 606 ISTINGPLVLLILLLLGPCIILNRLVQFVKDRISVVQALV-----LTQQY----- 650
QY 61 PRSDVNDIKGTPPEISAAQ 80
Db 651 -----HQLKSIDPEMESRE 665
Search completed: April 19, 2004, 10:53:35
Job time : 17 secs

range 457-538

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:41:14 ; Search time 39 Seconds

(without alignments)
736.209 Million cell updates/sec

Title: US-10-069-883-1_COPY_448_538

Perfect score: 457

Sequence: 1 ILPFLGLAAIILLLFGPC.....PPEISAAQPLLRPSAGSS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 457 | 100.0 | 242 | 4 Q95245 | Q95245 homo sapien |
| 2 | 457 | 100.0 | 538 | 4 Q9NZG3 | Q9NZG3 homo sapien |
| 3 | 457 | 100.0 | 538 | 4 Q9UQF0 | Q9UQF0 homo sapien |
| 4 | 457 | 100.0 | 538 | 4 Q9NRZ2 | Q9NRZ2 homo sapien |
| 5 | 435 | 95.2 | 533 | 4 Q9NH77 | Q9NH77 homo sapien |
| 6 | 353 | 77.2 | 542 | 15 Q991W9 | Q991W9 multiple sc |
| 7 | 267 | 58.4 | 52 | 4 Q9UQF1 | Q9UQF1 homo sapien |
| 8 | 258 | 56.5 | 129 | 4 Q8NHW1 | Q8NHW1 homo sapien |
| 9 | 141 | 30.9 | 144 | 4 Q8NC12 | Q8NC12 homo sapien |
| 10 | 117 | 25.6 | 555 | 15 Q9N2J8 | Q9N2J8 herv-h/env6 |
| 11 | 114 | 24.9 | 584 | 4 Q9UNM3 | Q9UNM3 herv-h/env6 |
| 12 | 114 | 24.9 | 584 | 15 Q9N2K0 | Q9N2K0 herv-h/env6 |
| 13 | 113 | 24.7 | 665 | 15 Q67649 | Q67649 gibbon ape |
| 14 | 113 | 24.7 | 668 | 15 Q9YWL9 | Q9YWL9 simian sarc |
| 15 | 112.5 | 24.6 | 620 | 6 Q9GLF7 | Q9GLF7 trichosurus |
| 16 | 112 | 24.5 | 666 | 15 Q90AE9 | Q90AE9 feline leuk |

Q85510 xenotropic
Q9d22 murine leuk
Q90r13 murine leuk
Q8um29 moloney mur
Q04596 mus musculu
Q61919 mus musculu
P70356 mus musculu
Q7svk6 murine leuk
Q83399 murine leuk
Q9yy83 murine leuk
Q7slw4 murine leuk

ALIGNMENTS

RESULT 1
ID Q95245 PRELIMINARY; PRT; 242 AA.
AC Q95245;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99039005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.,
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072508; AAD14548.1; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 1
SQ SEQUENCE 242 AA; 26857 MW; A03512693D346F9 CRC64;
Query Match 100.0%; Score 457; DB 4; Length 242;
Best Local Similarity 100.0%; Pred. No. 2.3e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ILPLGLAAIILLGLPGCIFNLLVNFVSRIRIAVAVKLMQEPKQSKTKIYRPLDRPAS 60
Db 152 ILPLGLAAIILLGLPGCIFNLLVNFVSRIRIAVAVKLMQEPKQSKTKIYRPLDRPAS 211
QY 61 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
Db 212 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 242
RESULT 2
Q9NZG3 PRELIMINARY; PRT; 538 AA.
ID Q9NZG3
AC Q9NZG3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Syncytin precursor (Enverin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

TISSUE=Testis;
MEDLINE=20155476; PubMed=10693809;
Sna M., Lee X., Li X., Veldman G.M., Finnerty H., Racie L.,
Lavallie E., Tang X., Edouard P., Howes S., Keith J.C. Jr.,
McCoy J.M.;
"Syncytin is a captive retroviral envelope protein involved in human
placental morphogenesis.";
Nature 403:785-789(2000).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99039005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.,
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072506; AAD14546.2; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
QY 61 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
Db 508 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 538
RESULT 3
Q9UCFO PRELIMINARY; PRT; 538 AA.
ID Q9UCFO;
AC Q9UCFO;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Envelope protein precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99039005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.,
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family.";
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072506; AAD14546.2; -
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW SIGNAL.
FT CHAIN 1 21 POTENTIAL.
FT CHAIN 22 538 ENVELOPE PROTEIN.
SQ SEQUENCE 538 AA; 59866 MW; C54648A3C7043870 CRC64;
Query Match 100.0%; Score 457; DB 4; Length 538;
Best Local Similarity 100.0%; Pred. No. 5.7e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 60
DB 448 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 507
QY 61 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
DB 508 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 538
RESULT 4
ID Q9NRZ2 PRELIMINARY; PRT; 538 AA.
AC Q9NRZ2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope protein.
GN ENV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20284713; PubMed=10825480;
RA Voisset C., Bouton O., Bedin F., Duret L., Mandrand B., Mallet F.,
Paranhos-Baccala G.,
RT "Chromosomal distribution and coding capacity of the human endogenous
retrovirus HERV-W family.";
RL AIDS Res. Hum. Retroviruses 16:731-740(2000).
DR EMBL; AF156963; AAF74215.1; -.
DR HSSP; P03385; IMOF.
DR Genew; HGNC:13525; ERVWE1.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
SQ SEQUENCE 538 AA; 59894 MW; 476998A13A2A256F CRC64;
Query Match 100.0%; Score 457; DB 4; Length 538;
Best Local Similarity 100.0%; Pred. No. 5.7e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 60
DB 448 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 507
QY 61 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
DB 508 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 538
RESULT 5
ID Q8NHV7 PRELIMINARY; PRT; 533 AA.
AC Q8NHV7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Enverin (Fragment).
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21985840; PubMed=11990458;
RA Alliel P.M., Perin J.P., Pierig R., Nusebaum J.L., Menard A.,
Rieger F.;
RT "Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and
its env transcripts.";
RL C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21985840; PubMed=11990458;
RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.;
RT "The HERV-W/7q family in the human genome. Potential for protein
expression and gene regulation.";
RL Cell. Mol. Biol. 48:213-217(2002).
DR EMBL; AF506835; AAM33413.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
FT NON TER 533 -533
SQ SEQUENCE 533 AA; 59503 MW; 14FB789BCE1C41CD CRC64;
Query Match 95.2%; Score 435; DB 4; Length 533;
Best Local Similarity 100.0%; Pred. No. 2.1e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 60
DB 448 ILPFLGPLAAIILLFPGCIENLVNFVSSRIEAVKLMQEPKMQSKTKIYRPLDRPAS 507
QY 61 PRSDVNDIKGTPPEISAAQPLLRN 86
DB 508 PRSDVNDIKGTPPEISAAQPLLRN 533
RESULT 6
Q991W9 PRELIMINARY; PRT; 542 AA.
ID Q991W9;
AC Q991W9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Recombinant envelope protein.
GN ENV.
OS Multiple sclerosis associated retrovirus element.
OC Viruses; Retroviridae; Retroviridae.
OX NCBI_TaxID=89382;
RN [1]
RP SEQUENCE FROM N.A.
RA Perron H., Jouvin-Marche E., Michel M., Ounanian-Paraz A., Camelo S.,
Dumon A., Jolivet-Reynaud C., Marcel F., Souillet Y., Borel E.,
Gebuhrer L.;
RT "Multiple sclerosis retrovirus-induced polyclonal Vb16 T-cell receptor
activation as a trigger of abnormal immune response.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF331500; AAK18189.1; -.
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Envelope protein.
SQ SEQUENCE 542 AA; 60625 MW; AAF3A8322E03C9EE CRC64;
Query Match 77.2%; Score 353; DB 15; Length 542;
Best Local Similarity 79.8%; Pred. No. 9e-33;

Matches 75; Conservative 2; Mismatches 13; Indels 4; Gaps 1;
QY 2 LPFLGFLAAILLLIFGCIENLLVNFVSSRIEAVK----LQMEPKQSKTKIYRRPLDR 57
Db 449 LPFLGFLAAILLLIFGCIENFLVFKVSSRIEAVKQIVLQMEFQWQSKTKIYRGPLDR 509
QY 58 PASPSVDNDIKGTPPEISAAQPLLRNSAGSS 91
Db 509 PARLCSVDNDIEVTPPEISAAQPLLRNSVGS 542
RESULT 7
Q9UQF1 PRELIMINARY; PRT; 52 AA.
ID Q9UQF1
AC Q9UQF1
DT 01-OCT-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Envelope protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99099005; PubMed=9882319;
RA Blond J.L., Beseme F., Duret L., Bouton O., Bedin F., Perron H.,
RA Mandrand B., Mallet F.,
RT "Molecular characterization and placental expression of HERV-W, a new
RT human endogenous retrovirus family."
RL J. Virol. 73:1175-1185(1999).
DR EMBL; AF072507; AAD14547.1;
FT NON TER 1
SQ SEQUENCE 52 AA; 5719 MW; 2BDD668523DCB4F4 CRC64;
Query Match 58.4%; Score 267; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 7.6e-24; Indels 0; Gaps 0;
Matches 52; Conservative 0; Mismatches 0;
QY 40 MEFPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
Db 1 MEFPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 52
RESULT 8
Q8NHW1 PRELIMINARY; PRT; 129 AA.
ID Q8NHW1
AC Q8NHW1
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Everin-2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=99052087; PubMed=9835022;
RA Alliel P.M., Perin J.P., Pierig R., Nussbaum J.L., Menard A.,
RA Rieger F.,
RT "Endogenous retroviruses and multiple sclerosis. II. HERV-7q."
RL C. R. Acad. Sci. III, Sci. Vie 321:857-863(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=21985839; PubMed=11990457;
RA Alliel P.M., Perin J.P., Goudou D., Bitoun M., Robert B., Rieger F.,
RA "The HERV-W/7q family in the human genome. Potential for protein
RT expression and gene regulation."
RL Cell. Mol. Biol. 48:205-212(2002).

EMBL; AY098482; AAM33141.1; --
FT NON TER 1
FT NON TER 129 129
SQ SEQUENCE 129 AA; 14217 MW; FF319F08338DE84 CRC64;
Query Match 56.5%; Score 258; DB 4; Length 129;
Best Local Similarity 73.0%; Pred. No. 2.4e-22;
Matches 54; Conservative 4; Mismatches 12; Indels 4; Gaps 1;
QY 17 FGPCIFENLLVNFVSSRIEAVK---LQMEPKQSKTKIYRRPLDRPASPRSDVNDIKGTP 72
Db 56 FNPCKFLVFKVSSRIEAVKQIVLQMEFQWQSKTKIYCRPLDRPASPRSDVNDIKGTP 115
QY 73 PEISAAQPLLRN 86
Db 116 PEISTARPLSCP 129
RESULT 9
Q8NC12 PRELIMINARY; PRT; 144 AA.
ID Q8NC12
AC Q8NC12
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein FLJ90611.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.,
RT "NEDO human cDNA sequencing project."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK075092; BAC11396.1;
DR GO; GO:0019028; C: viral capsid; IEA.
DR GO; GO:0005198; F: structural molecule activity; IEA.
DR InterPro; IPR002050; Env polyprotein.
DR Pfam; PF00429; Env polyprotein; 1.
KW Hypothetical protein.
SQ SEQUENCE 144 AA; 16165 MW; 05BC2EE7BF1905A4 CRC64;
Query Match 30.9%; Score 141; DB 4; Length 144;
Best Local Similarity 61.5%; Pred. No. 1.5e-08;
Matches 24; Conservative 9; Mismatches 6; Indels 0; Gaps 0;
QY 1 ILPFLGFLAAILLLIFGCIENLLVNFVSSRIEAVKQ 39
Db 89 VLPTGFLVSLLLLLFGPCLLNLTQFVSSRLQAIKQ 127
RESULT 10
Q9N2J8 PRELIMINARY; PRT; 555 AA.
ID Q9N2J8
AC Q9N2J8
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Envelope protein.
GN ENV.
OS HERV-H/env62.
OC Viruses; Retroviridae.
OX NCBI_TaxID=129525;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21109114; PubMed=11162811;
RA de Parseval N., Casella J.F., Gressin L., Heidmann T.

RT "Characterization of the three envelopes of the HERV-H family
encompassing the ISU domain and evolutionary history in primates.";
RL Virology 279:558-569 (2001).
DR EMBL; AJ289711; CAB94194.1; -.
DR HSSP; P03385; IMOF.

DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
SQ SEQUENCE 555 AA; 60910 MW; 8559604545407844 CRC64;

Query Match 25.6%; Score 117; DB 15; Length 555;
Best Local Similarity 58.3%; Pred. No. 4.4e-05;
Matches 21; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 ILPLGGLAAIILLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 58
Db 515 VLPISPLIPILFLLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 582

RESULT 11

Q9UNM3 PRELIMINARY; PRT; 584 AA.
AC Q9UNM3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Env protein.
GN Env.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99296739; PubMed=10366582;
RA Lindeskog M., Mager D.L., Blomberg J.;
RT "Isolation of a human endogenous retroviral HERV-H element with an
open env reading frame.";
RL Virology 258:441-450 (1999).
DR EMBL; AF108843; AAD34324.1; -.
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 2.
SQ SEQUENCE 584 AA; 64298 MW; C841C55A3624FD0A CRC64;

Query Match 24.9%; Score 114; DB 4; Length 584;
Best Local Similarity 41.9%; Pred. No. 0.00011;
Matches 26; Conservative 10; Mismatches 24; Indels 2; Gaps 1;

QY 1 ILPLGGLAAIILLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 58
Db 523 VLPISPLIPILFLLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 582

QY 59 AS 60

Db 583 SA 584

RESULT 12

Q9N2K0 PRELIMINARY; PRT; 584 AA.
AC Q9N2K0
DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Envelope protein.
GN Env.
OS HERV-H/env62.
OC Viruses; Retroviral viruses; Retroviridae.
OX NCBI_TaxID=129525;

[1]

RN SEQUENCE FROM N.A.
RX MEDLINE=21109114; PubMed=11162811;
RA de Parreval N., Casella J.F., Gressin L., Heidmann T.;
RT "Characterization of the three envelopes of the HERV-H family
encompassing the ISU domain and evolutionary history in primates.";
RL Virology 279:558-569 (2001).
DR EMBL; AJ289709; CAB94192.1; -.
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 2.
SQ SEQUENCE 584 AA; 64318 MW; C03D260B5A60BDAB CRC64;

Query Match 24.9%; Score 114; DB 15; Length 584;
Best Local Similarity 41.9%; Pred. No. 0.00011;
Matches 26; Conservative 10; Mismatches 24; Indels 2; Gaps 1;

QY 1 ILPLGGLAAIILLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 58
Db 523 VLPISPLIPILFLLFGPCIFNLVNFVSRIEAVKQMEPKMOSKTKIYRPL--DRP 582

QY 59 AS 60

Db 583 SA 584

RESULT 13

Q67649 PRELIMINARY; PRT; 665 AA.
AC Q67649
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Envelope protein.
OS Gibbon ape leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11840;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson C.A., Changgao Y., Hooker D., Deacon N.J., Eiden M.V.;
RT "Chinese hamster cells restrict entry of gibbon ape leukemia
viruses.";
RL Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U20589; AAA69474.1; -.
DR HSSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
SQ SEQUENCE 665 AA; 73832 MW; 1005BB3559310405 CRC64;

Query Match 24.7%; Score 113; DB 15; Length 665;
Best Local Similarity 48.8%; Pred. No. 0.00016;
Matches 20; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 6 GPLAAIILLFGPCIFNLVNFVSRIEAVKQMEPKMOS 46
Db 618 GPLLLLLIILGPCIINELVQINDRVSAVKILLRQKYQT 658

RESULT 14

Q9YWL9 PRELIMINARY; PRT; 668 AA.
AC Q9YWL9
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Envelope protein.
OS Simian sarcoma-associated virus.
OC Viruses; Retroviral viruses; Retroviridae; Mammalian type C retroviruses.


```
RESULT 18
Q85518 PRELIMINARY; PRT; 642 AA.
AC Q85518; Q85520; (Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Envelope protein precursor.
OS Feline leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86119207; PubMed=2828667;
RA Donahue P.R., Hoover E.A., Beltz G.A., Riedel N., Hirsch V.M.,
RA Overbaugh J., Mullins J.I.;
RT "Strong sequence conservation among horizontally transmissible,
RT minimally pathogenic feline leukemia viruses.";
RL J. Virol. 62:722-731(1988).
DR EMBL; M18248; AAA43051.1; -.
DR FIK; B46165; B46165.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; Env_Polyprotein; 1.
DR SIGNAL.
KW SIGNAL.
FT CHAIN 1 33 POTENTIAL.
FT CHAIN 34 444 POTENTIAL.
FT CHAIN 446 642 POTENTIAL.
SQ SEQUENCE 642 AA; 71080 MW; DC4AA8F9ED8A574E CRC64;

Query Match 24.1%; Score 110; DB 15; Length 642;
Best Local Similarity 43.1%; Pred. No. 0.00035;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGPLAAIILLFGPCIFNLLVNFVSRIEAVKLMQPKMQSKTKIYRPLDRP 58
DB 587 ISSIMGPLLILLILLFGPCILNLVQFKDRISVQALILTQYQYQIKQY--DPRP 642

RESULT 19
Q85522 PRELIMINARY; PRT; 642 AA.
AC Q85522;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Env gene polyprotein precursor.
OS Feline leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Felv-FAIDS;
RA Donahue P.R., Hoover E.A., Beltz G.A., Riedel N., Hirsch V.M.,
RA Overbaugh J., Mullins J.I.;
RT "Strong sequence conservation among horizontally transmissible,
RT minimally pathogenic feline leukemia viruses.";
RL J. Virol. 62:722-731(1988).
DR EMBL; M18247; AAA93093.1; -.
DR FIK; T10533; T10533.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; Env_Polyprotein; 1.
DR Polyprotein; Signal.
KW Polyprotein; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 445 POTENTIAL.
```

```
FT CHAIN 446 642 POTENTIAL.
SQ SEQUENCE 642 AA; 71081 MW; FDS8B54E6FD2AA5D CRC64;

Query Match 24.1%; Score 110; DB 15; Length 642;
Best Local Similarity 43.1%; Pred. No. 0.00035;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGPLAAIILLFGPCIFNLLVNFVSRIEAVKLMQPKMQSKTKIYRPLDRP 58
DB 587 ISSIMGPLLILLILLFGPCILNLVQFKDRISVQALILTQYQYQIKQY--DPRP 642

RESULT 20
Q85812 PRELIMINARY; PRT; 642 AA.
AC Q85812;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Env gene polyprotein.
GN ENV
OS Feline leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rickard subgroup A;
RX MEDLINE=98362106; PubMed=9696797;
RX Chen H., Bechtel M.K., Shi Y., Phipps A., Mathes L.E., Hayes K.A.,
RX Roy-Burman P.;
RT "Pathogenicity induced by feline leukemia virus, rickard strain,
RT subgroup A plasmid DNA (pFRA).";
RL J. Virol. 72:7048-7056(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Rickard subgroup A;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF052723; AAC31802.1; -.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_Polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; Env_Polyprotein; 1.
DR Polyprotein.
KW CHAIN 34 445 GP70 PROTEIN.
FT CHAIN 446 642 P15E PROTEIN.
SQ SEQUENCE 642 AA; 70962 MW; 11CE7A13BBAF7C22 CRC64;

Query Match 24.1%; Score 110; DB 15; Length 642;
Best Local Similarity 43.1%; Pred. No. 0.00035;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 1 ILPLGPLAAIILLFGPCIFNLLVNFVSRIEAVKLMQPKMQSKTKIYRPLDRP 58
DB 587 ISSIMGPLLILLILLFGPCILNLVQFKDRISVQALILTQYQYQIKQY--DPRP 642

RESULT 21
Q85525 PRELIMINARY; PRT; 642 AA.
AC Q85525;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Envelope polyprotein precursor.
OS Feline leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RA Donahue P.R.;
```

```

RL Submitted (OCT-1988) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88127161; PubMed=2893454;
RA Overbaugh J., Donahue P.R., Quackenbush S.L., Hoover E.A.,
RA Mullins J.I.;
RT "Molecular cloning of a feline leukemia virus that induces fatal
RT immunodeficiency disease in cats.";
RL Science 239:906-910(1988).
DR EMBL; M18246; AAA43062.1; -.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Polyprotein; Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 445 ENVELOPE PROTEIN GP70.
FT CHAIN 446 642 ENVELOPE PROTEIN P15E.
SQ SEQUENCE 642 AA; 70924 MW; 45AABF65D95364D6 CRC64;
Query Match 24.1%; Score 110; DB 15; Length 642;
Best Local Similarity 43.1%; Pred. No. 0.00035;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;
QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVKQMEPKMQSKTKIYRRPLDRP 58
Db 587 ISSIMGPLLLILLLLFGPCILNRLVQFVKDRISVQALILQYQIQY--DPRP 642

RESULT 22
Q01441
ID O01441 PRELIMINARY; PRT; 646 AA.
AC O01441;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope polyprotein.
GN ENV.
OS Feline leukemia virus.
OC Viruses; Retroviridae; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98184498; PubMed=9525586;
RA Rohn J.L., Moser M.S., Gwynn S.R., Baldwin D.N., Overbaugh J.;
RT "In vivo evolution of a novel, syncytium-inducing and cytopathic
RT feline leukemia virus variant.";
RL J. Virol. 72:2686-2696(1998).
DR EMBL; U58951; AAC59315.1; -.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
KW Polyprotein.
SQ SEQUENCE 646 AA; 71371 MW; BBBDAF7052E4427 CRC64;
Query Match 24.1%; Score 110; DB 15; Length 646;
Best Local Similarity 43.1%; Pred. No. 0.00035;
Matches 25; Conservative 8; Mismatches 23; Indels 2; Gaps 1;
QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVKQMEPKMQSKTKIYRRPLDRP 58
Db 591 ISSIMGPLLLILLLLFGPCILNRLVQFVKDRISVQALILQYQIQY--DPRP 646

RESULT 23
Q01Y75
ID Q01Y75 PRELIMINARY; PRT; 671 AA.
AC Q01Y75;

```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Envelope.
GN ENV.
OS Mus musculus castaneus (Southeastern Asian house mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10091;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bgr; TISSUE=Liver;
RX MEDLINE=21232482; PubMed=1133885;
RA Ikeda H., Kato K., Kitani H., Suzuki T., Yoshida T., Inaguma Y.,
RA Yamamoto N., Suh J.G., Hyun B.H., Yamagata T., Namikawa T., Tomita T.;
RT "Virological properties and nucleotide sequences of Cas-E-type
RT Endogenous Ecotropic Murine Leukemia Viruses in South Asian Wild Mice,
RT Mus musculus castaneus.";
RL J. Virol. 75:5049-5058(2001).
DR EMBL; AB050721; BAB47170.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR InterPro; IPR008981; F_MuLV_bind.
DR Pfam; PF00429; ENV_polyprotein; 1.
SQ SEQUENCE 671 AA; 73473 MW; 6117F85263DC1A39 CRC64;
Query Match 23.9%; Score 109; DB 11; Length 671;
Best Local Similarity 32.5%; Pred. No. 0.00048;
Matches 26; Conservative 12; Mismatches 22; Indels 20; Gaps 2;
QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVKQMEPKMQSKTKIYRRPLDRPAS 60
Db 612 ISTIMGPLLLILLLLFGPCILNRLVQFVKDRISVQALI-----LTQY----- 656

QY 61 PRSDVNDIKGTPPREISAQAQ 80
Db 657 -----HQLKSTIEPEVESRE 671

RESULT 24
Q05626
ID Q05626 PRELIMINARY; PRT; 201 AA.
AC Q05626;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mink cell focus-forming 247 MuLV env (Fragment).
OS Murine leukemia virus.
OC Viruses; Retroviridae; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11786;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83112196; PubMed=6296437;
RA Kelly M., Holland C.A., Lung M.L., Chattopadhyay S.K., Lowy D.R.,
RA Hopkins N.H.;
RT "Nucleotide sequence of the 3' end of MCF 247 murine leukemia virus.";
RL J. Virol. 45:291-298(1983).
DR EMBL; J02249; AAA46513.1; -.
DR HSP; P03385; IMOF.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; ENV_polyprotein; 1.
FT NON_TER 1 1
FT CHAIN 181 201 POTENTIAL.
SQ SEQUENCE 201 AA; 22358 MW; 9338409B29EC9EBC CRC64;
Query Match 23.6%; Score 108; DB 15; Length 201;
Best Local Similarity 32.5%; Pred. No. 0.00016;
Matches 26; Conservative 12; Mismatches 22; Indels 20; Gaps 2;
QY 1 ILPLGLPLAAILLLFGPCIFNLLVNFVSSRIEAVKQMEPKMQSKTKIYRRPLDRPAS 60

```

Db 142 ISTINGPLIILLILLFGPCILNRLVQVKDRISVQALV-----LTQOY----- 186

QY 61 PRSDVNDIKGTPPEEISAAQ 80
Db 187 -----HQLKSIDPEVESRE 201

RESULT 25

Q8K170
ID Q8K170 PRELIMINARY; PRT; 325 AA.
AC Q8K170;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, last annotation update)
DE Hypothetical protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028259; AAH28259.1; .
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR002050; Env_polyprotein.
DR Pfam; PF00429; Env_polyprotein; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 325 AA; 35496 MW; 63DBAB1D734B5072 CRC64;

Query Match 23.6%; Score 108; DB 11; Length 325;
Best Local Similarity 32.5%; Fred. No. 0.00027;
Matches 26; Conservative 12; Mismatches 22; Indels 20; Gaps 2;

QY 1 ILPLFLGPLAAIILLFGPCILNRLVQVKDRISVQALV-----LTQOY----- 310
Db 266 ISTINGPLIILLILLFGPCILNRLVQVKDRISVQALV-----LTQOY----- 310

QY 61 PRSDVNDIKGTPPEEISAAQ 80
Db 311 -----HQLKSIDPEVESRE 325

Search completed: April 19, 2004, 10:54:26
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:52:09 ; Search time 22 Seconds
(without alignments)
213.544 Million cell updates/sec

Title: US-10-069-883-1_COPY_448_538

Perfect score: 457

Sequence: 1 ILPLGLPLAIIILLFPGPC.....PPREISAAQPLLRPNAGSS 91

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | ID | Description |
|------------|-------|-------------|-----|--------------------|
| 1 | 457 | 100.0 | 538 | US-09-175-928-4 |
| 2 | 353 | 77.2 | 162 | US-08-978-847B-109 |
| 3 | 287 | 62.8 | 77 | US-08-978-847B-103 |
| 4 | 204 | 44.6 | 433 | US-08-691-563C-87 |
| 5 | 204 | 44.6 | 433 | US-09-374-766-87 |
| 6 | 204 | 44.6 | 433 | US-08-978-847B-81 |
| 7 | 189 | 41.4 | 493 | US-08-978-847B-106 |
| 8 | 132.5 | 29.0 | 152 | US-09-120-653D-7 |
| 9 | 105 | 23.0 | 75 | US-09-042-012-8 |
| 10 | 105 | 23.0 | 75 | US-09-308-086-2 |
| 11 | 105 | 23.0 | 75 | US-09-457-324-8 |
| 12 | 105 | 23.0 | 645 | US-09-315-127-9 |
| 13 | 105 | 23.0 | 645 | US-09-315-127-8 |
| 14 | 105 | 23.0 | 654 | US-09-315-127-11 |
| 15 | 105 | 23.0 | 654 | US-09-315-127-12 |
| 16 | 103 | 22.5 | 192 | US-08-486-099-107 |
| 17 | 103 | 22.5 | 192 | US-08-360-107A-117 |
| 18 | 103 | 22.5 | 192 | US-08-484-223B-107 |
| 19 | 103 | 22.5 | 192 | US-08-919-597-107 |
| 20 | 103 | 22.5 | 192 | US-08-475-668A-107 |
| 21 | 103 | 22.5 | 192 | US-08-485-551A-107 |
| 22 | 103 | 22.5 | 192 | US-08-471-913A-107 |
| 23 | 103 | 22.5 | 192 | US-08-485-264A-107 |
| 24 | 103 | 22.5 | 192 | US-08-474-349A-107 |
| 25 | 103 | 22.5 | 192 | US-08-470-896-107 |
| 26 | 103 | 22.5 | 192 | US-08-485-546A-107 |
| 27 | 103 | 22.5 | 196 | US-08-484-126-7 |

| | | | | | |
|----------------------|------|------|------|-----|--------------------|
| US-09-374-909-7 | 196 | 22.5 | 103 | 28 | Sequence 7, Appli |
| US-09-315-127-2 | 632 | 22.5 | 103 | 29 | Sequence 2, Appli |
| US-09-315-127-3 | 632 | 22.5 | 103 | 30 | Sequence 3, Appli |
| US-09-309-572-14 | 665 | 22.5 | 103 | 31 | Sequence 14, Appl |
| US-09-718-096-14 | 665 | 22.5 | 103 | 32 | Sequence 14, Appl |
| US-09-554-572-26 | 1312 | 22.5 | 103 | 33 | Sequence 26, Appl |
| US-09-075-272-5 | 673 | 21.9 | 100 | 34 | Sequence 5, Appli |
| US-09-555-352-10 | 232 | 21.2 | 97 | 35 | Sequence 10, Appl |
| US-08-007-282B-2 | 567 | 21.2 | 97 | 36 | Sequence 2, Appli |
| US-09-376-781-6 | 638 | 21.2 | 97 | 37 | Sequence 6, Appli |
| US-09-315-127-5 | 667 | 20.6 | 94 | 38 | Sequence 5, Appli |
| US-09-315-127-6 | 667 | 20.6 | 94 | 39 | Sequence 6, Appli |
| US-09-376-781-3 | 656 | 19.7 | 90 | 40 | Sequence 3, Appli |
| US-09-111-085-4 | 657 | 19.7 | 90 | 41 | Sequence 4, Appli |
| US-09-376-781-4 | 657 | 19.7 | 90 | 42 | Sequence 2, Appli |
| US-09-111-085-2 | 660 | 19.7 | 90 | 43 | Sequence 5, Appli |
| US-09-376-781-5 | 660 | 19.7 | 90 | 44 | Sequence 5, Appli |
| US-09-489-039A-11606 | 404 | 15.2 | 69.5 | 45 | Sequence 11606, A |
| US-09-252-991A-28564 | 188 | 14.7 | 67 | 46 | Sequence 28564, A |
| US-09-543-681A-5915 | 533 | 14.6 | 66 | 47 | Sequence 5915, Ap |
| US-08-235-836C-68 | 693 | 14.4 | 66 | 48 | Sequence 68, Appl |
| US-08-871-572B-1 | 337 | 14.1 | 64.5 | 49 | Sequence 1, Appli |
| US-08-871-572B-4 | 337 | 14.1 | 64.5 | 50 | Sequence 4, Appli |
| US-09-265-540E-5 | 337 | 14.1 | 64.5 | 51 | Sequence 5, Appli |
| US-08-752-307B-11 | 612 | 13.8 | 63 | 52 | Sequence 11, Appl |
| US-09-707-802-11 | 612 | 13.8 | 63 | 53 | Sequence 11, Appl |
| US-09-991-326-11 | 612 | 13.8 | 63 | 54 | Sequence 11, Appl |
| US-09-220-641-3 | 1257 | 13.8 | 63 | 55 | Sequence 3, Appli |
| US-08-506-296B-28 | 1268 | 13.8 | 63 | 56 | Sequence 28, Appli |
| US-09-081-975-3 | 635 | 13.7 | 62.5 | 57 | Sequence 3, Appli |
| US-09-328-352-6216 | 955 | 13.7 | 62.5 | 58 | Sequence 6216, Ap |
| US-09-362-842-8 | 1153 | 13.7 | 62.5 | 59 | Sequence 8, Appli |
| US-09-252-991A-27306 | 511 | 13.6 | 62 | 60 | Sequence 27306, A |
| US-09-252-991A-17902 | 488 | 13.5 | 61 | 61 | Sequence 17902, A |
| US-09-252-991A-29395 | 759 | 13.5 | 61.5 | 62 | Sequence 29395, A |
| US-09-198-452A-1139 | 288 | 13.3 | 61 | 63 | Sequence 1139, Ap |
| US-09-328-352-7154 | 377 | 13.2 | 60.5 | 64 | Sequence 7154, Ap |
| US-09-252-991A-24651 | 256 | 13.1 | 60 | 65 | Sequence 24651, A |
| US-09-286-981B-7 | 428 | 13.1 | 60 | 66 | Sequence 7, Appli |
| US-09-562-737-19 | 659 | 13.1 | 67 | 67 | Sequence 19, Appl |
| US-08-235-836C-72 | 693 | 13.1 | 60 | 68 | Sequence 72, Appl |
| US-09-252-991A-26790 | 172 | 13.0 | 59.5 | 69 | Sequence 26790, A |
| US-09-252-991A-26342 | 336 | 13.0 | 59.5 | 70 | Sequence 26342, A |
| US-09-198-452A-521 | 369 | 13.0 | 59.5 | 71 | Sequence 521, App |
| US-08-480-547A-23 | 875 | 13.0 | 59.5 | 72 | Sequence 23, Appl |
| US-08-250-847B-23 | 875 | 13.0 | 59.5 | 73 | Sequence 23, Appl |
| US-08-463-949A-23 | 875 | 13.0 | 59.5 | 74 | Sequence 23, Appl |
| US-08-464-410A-23 | 875 | 13.0 | 59.5 | 75 | Sequence 23, Appl |
| US-09-226-741-5 | 875 | 13.0 | 59.5 | 76 | Sequence 5, Appli |
| US-09-595-514-5 | 875 | 13.0 | 59.5 | 77 | Sequence 5, Appli |
| PCT-US94-06066-23 | 3079 | 13.0 | 59.5 | 78 | Sequence 23, Appl |
| US-09-413-814-80 | 100 | 12.9 | 59 | 79 | Sequence 80, Appl |
| US-09-152-060-61 | 106 | 12.9 | 59 | 80 | Sequence 61, Appl |
| US-09-663-600A-189 | 106 | 12.9 | 59 | 81 | Sequence 189, App |
| US-09-252-991A-29392 | 487 | 12.9 | 59 | 82 | Sequence 29392, A |
| US-09-328-352-4521 | 758 | 12.9 | 59 | 83 | Sequence 4521, Ap |
| US-09-134-001C-4917 | 758 | 12.9 | 59 | 84 | Sequence 4917, Ap |
| US-09-352-991A-23920 | 146 | 12.8 | 58.5 | 85 | Sequence 23920, A |
| US-08-118-270-16 | 362 | 12.8 | 58.5 | 86 | Sequence 16, Appl |
| PCT-US93-08528-16 | 1128 | 12.8 | 58.5 | 87 | Sequence 16, Appl |
| US-09-328-352-4973 | 1709 | 12.8 | 58.5 | 88 | Sequence 4973, Ap |
| US-09-392-812A-6 | 196 | 12.7 | 58 | 89 | Sequence 6, Appli |
| US-08-336-728A-44 | 208 | 12.7 | 58 | 90 | Sequence 44, Appl |
| US-08-836-252A-6 | 208 | 12.7 | 58 | 91 | Sequence 6, Appli |
| US-08-836-252A-46 | 208 | 12.7 | 58 | 92 | Sequence 46, Appl |
| US-09-224-681-46 | 208 | 12.7 | 58 | 93 | Sequence 46, Appl |
| US-08-336-728A-46 | 208 | 12.7 | 58 | 94 | Sequence 46, Appl |
| US-08-955-848A-82 | 248 | 12.7 | 58 | 95 | Sequence 82, Appl |
| US-08-220-379B-2 | 273 | 12.7 | 58 | 96 | Sequence 2, Appli |
| US-08-628-428-9 | 273 | 12.7 | 58 | 97 | Sequence 9, Appli |
| US-08-482-918-48 | 273 | 12.7 | 58 | 98 | Sequence 48, Appl |
| US-08-482-918-49 | 273 | 12.7 | 58 | 99 | Sequence 49, Appl |
| US-08-482-918-50 | 273 | 12.7 | 58 | 100 | Sequence 50, Appl |

ALIGNMENTS

```
RESULT 1
US-09-175-928-4
; Sequence 4, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 606B.AJ172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-175-928-4

Query Match          100.0%; Score 457; DB 4; Length 538;
Best Local Similarity 100.0%; Pred. No. 9.7e-48;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPFLGPLAAIILLFPGPCIFNLVNFVSSRIEAVKQLQVLMQEPKQSKTIYRPLDRPAS 60
DB 448 ILPFLGPLAAIILLFPGPCIFNLVNFVSSRIEAVKQLQVLMQEPKQSKTIYRPLDRPAS 507

QY 61 PRSDVNDIKGTPPEEISAAQPLLRNSAGSS 91
DB 508 PRSDVNDIKGTPPEEISAAQPLLRNSAGSS 538

RESULT 2
US-08-979-847B-109
; Sequence 109, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKER, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-98
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A

US-08-979-847B-109
; Sequence 109, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKER, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-98
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-98
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 109:
US-08-979-847B-109

Query Match          77.2%; Score 353; DB 4; Length 162;
Best Local Similarity 79.8%; Pred. No. 1.2e-35;
Matches 75; Conservative 2; Mismatches 13; Indels 4; Gaps 1;

QY 2 LPFLGPLAAIILLFPGPCIFNLVNFVSSRIEAVKQLQVLMQEPKQSKTIYRPLDR 57
DB 69 LPFLGPLAAIILLFPGPCIFNLVNFVSSRIEAVKQLQVLMQEPKQSKTIYRPLDR 128

QY 58 PRSDVNDIKGTPPEEISAAQPLLRNSAGSS 91
DB 129 PARLCSVDNDIEVTPEEISTAQPLLRNSVSGSS 162

RESULT 3
US-08-979-847B-103
; Sequence 103, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKER, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
; THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-NOV-98
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 77 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 103:
US-08-979-847B-103

Query Match 62.8%; Score 287; DB 4; Length 77;
Best Local Similarity 77.9%; Pred. No. 5.7e-28;
Matches 60; Conservative 5; Mismatches 8; Indels 4; Gaps 1;
QY 19 PCIFNLVNFVSSRIEAVK-----LQMEPKMQSKTKIYRRPLDRPASPDRSDVDIKETPPE 74
Db 1 PCIFNLVNFVSSRIKTKLQIVLQMEHOMESMTKIHRRGLDRPASPDSVDNDIEGTPPE 60
QY 75 EISAQPLLRPNAGSS 91
Db 61 EISTAQPLLCNSAGSS 77

RESULT 4
US-08-691-563C-87
; Sequence 87, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 02-AUG-1996
; APPLICATION NUMBER: US/08/691,563C
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-2787
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-691-563C-87

Query Match 44.6%; Score 204; DB 3; Length 433;

Best Local Similarity 67.2%; Pred. No. 8.7e-17;
Matches 41; Conservative 7; Mismatches 9; Indels 4; Gaps 1;
QY 1 ILPFLGPLAAIILLLFGPCIFNLVNFVSSRIEAVK-----LQMEPKMQSKTKIYRRPLD 56
Db 118 LLPFLGPLMAIILLLTFGPCIFKLVKRVSSRIEAIKLVQVLMQVLPQMSSTNNFYGGPLE 177
QY 57 R 57
Db 178 R 178

RESULT 5
US-09-374-766-87
; Sequence 87, Application US/09374766
; Patent No. 6579526
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BEDIN
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 02-AUG-1996
; APPLICATION NUMBER: US/09/374,766
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 87:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-374-766-87

Query Match 44.6%; Score 204; DB 4; Length 433;
Best Local Similarity 67.2%; Pred. No. 8.7e-17;
Matches 41; Conservative 7; Mismatches 9; Indels 4; Gaps 1;
QY 1 ILPFLGPLAAIILLLFGPCIFNLVNFVSSRIEAVK-----LQMEPKMQSKTKIYRRPLD 56
Db 118 LLPFLGPLMAIILLLTFGPCIFKLVKRVSSRIEAIKLVQVLMQVLPQMSSTNNFYGGPLE 177
QY 57 R 57
Db 178 R 178

RESULT 6

US-08-979-847B-81
; Sequence 81, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC
; PARANHOS-BACCALA, GLAUCIA
; KOMURIAN-PRADEL, FLORENCE
; JOLIVET-REYNAUD, COLETTE
; MANDRAND, BERNARD
; GARSON, JEREMY
; TUBE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 81:

SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 81:
US-08-979-847B-81

Query Match 44.6%; Score 204; DB 4; Length 433;

Best Local Similarity 67.2%; Pred. No. 8.7e-17;
Matches 41; Conservative 7; Mismatches 9; Indels 4; Gaps 1;

QY 1 ILPFLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVK----LQMEPKMQSKTKIYRRPLD 56

Db 118 LLPLFLGPMIAILLLLFGPCIFKLVKFSVSRIRIAIKLQWLQMEFQMSSTNNFYQGPLE 177

QY 57 R 57

Db 178 R 178

RESULT 7

US-08-979-847B-106
; Sequence 106, Application US/08979847B
; Patent No. 6582703
; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
; BESEME, FREDERIC
; BEDIN, FREDERIC

PARANHOS-BACCALA, GLAUCIA
KOMURIAN-PRADEL, FLORENCE
JOLIVET-REYNAUD, COLETTE
MANDRAND, BERNARD
GARSON, JEREMY
TUBE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL
THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847B
FILING DATE: 26-No. 6582703-1997
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-08-979-847B-106

Query Match

Best Local Similarity 41.4%; Score 189; DB 4; Length 493;
Matches 38; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 ILPFLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKMQ 45

Db 448 VLPLFLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQVLQME 492

RESULT 8

US-09-120-653D-7
; Sequence 7, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:

APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL

TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
FILE REFERENCE: 98-338
CURRENT APPLICATION NUMBER: US/09/120,653D

CURRENT FILING DATE: 1998-07-22

PRIOR APPLICATION NUMBER: KR 98-10108

PRIOR FILING DATE: 1998-03-24

NUMBER OF SEQ ID NOS: 33

SOFTWARE: KopatentIn 1.71

SEQ ID NO 7

LENGTH: 152

; TYPE: PRT
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
; FEATURE:
; NAME/KEY: NON CONS
; LOCATION: (33)
; OTHER INFORMATION: any one of amino acids
; FEATURE:
; NAME/KEY: NON CONS
; LOCATION: (64)
; OTHER INFORMATION: any one of amino acids
US-09-120-653D-7

Query Match 29.0%; Score 132.5; DB 4; Length 152;
Best Local Similarity 51.7%; Pred. No. 1.3e-08;
Matches 31; Conservative 9; Mismatches 15; Indels 5; Gaps 2;
QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 56
Db 94 LLTGLGPVATVLLVLLGPGSIFNVLVKFV-SRFEAFKQVLMVLIKPRIISTNYFYGGHLD 152

RESULT 9
US-09-042-012-8
; Sequence 8, Application US/09042012A
; Patent No. 6111987
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042.012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Wildtype MuLV protein
US-09-042-012-8

Query Match 23.0%; Score 105; DB 3; Length 75;
Best Local Similarity 54.1%; Pred. No. 1.2e-05;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 21 ISTMGPLVILLILLFGPCILNRLVQFVKDRISVVQ 57

RESULT 10
US-09-305-086-2
; Sequence 2, Application US/09305086
; Patent No. 6150138
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-010
; CURRENT APPLICATION NUMBER: US/09/305.086
; CURRENT FILING DATE: 1999-05-04
; EARLIER APPLICATION NUMBER: US 08/816,439
; EARLIER FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Wildtype MuLV protein
US-09-305-086-2

Query Match 23.0%; Score 105; DB 3; Length 75;
Best Local Similarity 54.1%; Pred. No. 1.2e-05;

Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 21 ISTMGPLVILLILLFGPCILNRLVQFVKDRISVVQ 57
RESULT 11
US-09-457-324-8
; Sequence 8, Application US/09457324
; Patent No. 6277601
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/457.324
; CURRENT FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/042.012
; PRIOR FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Wildtype MuLV protein
US-09-457-324-8

Query Match 23.0%; Score 105; DB 3; Length 75;
Best Local Similarity 54.1%; Pred. No. 1.2e-05;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 21 ISTMGPLVILLILLFGPCILNRLVQFVKDRISVVQ 57

RESULT 12
US-09-315-127-8
; Sequence 8, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and
; TITLE OF INVENTION: Liposome Vectors and Use in Gene and Drug Therapy
; FILE REFERENCE: 44137-5023, U. of Tennessee
; CURRENT APPLICATION NUMBER: US/09/315.127
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-315-127-8

Query Match 23.0%; Score 105; DB 4; Length 645;
Best Local Similarity 54.1%; Pred. No. 0.00021;
Matches 20; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
QY 1 ILPFLGPLAAIILLFGPCIFNLLVNFVSSRIEAVK 37
Db 591 ISTMGPLVILLILLFGPCILNRLVQFVKDRISVVQ 627

RESULT 13
US-09-315-127-9
; Sequence 9, Application US/09315127
; Patent No. 6448390
; GENERAL INFORMATION:
; APPLICANT: The University of Tennessee, c/o Richard Cox
; TITLE OF INVENTION: Stable Envelope Proteins for Retroviral, Viral and

RESULT 17
US-08-360-107A-117
: Sequence 117, Application US/08360107A

```

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,223B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-029
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-223B-107

Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLFTGLAAIILLFGPCIEFLVNFVSSRIEAVK 37
DB 132 VMLPLGPLLCLLVLSFGPIIFNKMLTFIKHQIESIQ 168

RESULT 19
US-08-919-597-107
Sequence 107, Application US/08919597
Patent No. 6054265
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Patteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/919,597
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 08/470,896
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-919-597-107

Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLGFLAAIILLGPGCIENLLVNFVSSRIEAVK 37
DB 132 VMPLGLCLLLVLSFGPIFNKLMTFIKHQIESIQ 168

RESULT 20
US-08-475-668A-107
Sequence 107, Application US/08475668A
Patent No. 6060065
GENERAL INFORMATION:
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,668A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-475-668A-107

Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLGFLAAIILLGPGCIENLLVNFVSSRIEAVK 37
DB 132 VMPLGLCLLLVLSFGPIFNKLMTFIKHQIESIQ 168

RESULT 21
US-08-485-551A-107
Sequence 107, Application US/08485551A
Patent No. 6068973
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 211
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,551A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-023
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-551A-107

Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 ILPLGFLAAIILLGPGCIENLLVNFVSSRIEAVK 37
DB 132 VMPLGLCLLLVLSFGPIFNKLMTFIKHQIESIQ 168

RESULT 22
US-08-471-913A-107
Sequence 107, Application US/08471913A
Patent No. 6093794
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.

```

APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
NUMBER OF SEQUENCES: 214
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,913A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-030
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-471-913A-107
Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY      1 ILPLGLPAAITLLILFGPCIFNLVNFVSSRIEAVK 37
          :|:||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
DB      132 VMPLGLCLLVLVSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 24
US-08-474-349A-107
Sequence 107, Application US/08474349A
Patent No. 633395
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING
VIRUS TRANSMISSION
NUMBER OF SEQUENCES: 517
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,349A
FILING DATE: 07-JUN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-024
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 192 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-264A-107
Query Match 22.5%; Score 103; DB 3; Length 192;
Best Local Similarity 45.9%; Pred. No. 7.6e-05;
Matches 17; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY      1 ILPLGLPAAITLLILFGPCIFNLVNFVSSRIEAVK 37
          :|:||||:|:|:|:|:|:|:|:|:~
DB      132 VMPLGLCLLVLVSFGPIIFNKLMTFIKHQIESIQ 168

RESULT 23
US-08-485-264A-107
Sequence 107, Application US/08485264A
Patent No. 6228983
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING
RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
NUMBER OF SEQUENCES: 232
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

```



```

Matches      17; Conservative    11; Mismatches     9; Indels       0; Gaps       0;
Qy          1 ILPLGFLAIIILLPGFCIFNLVNFVSSRIEAVK 37
             ::|||:::|||||::|||::|||::|||::|||::
Db         132 VMPLGLPCLLLVLFGPIIFNKLMTFIKHOIESIQ 168

Search completed: April 19, 2004, 10:55:34
Job time : 24 secs
```

Search completed: April 19, 2004, 10:55:34
Job time : 24 secs

RESULT 25
US-08-470-896-107
; Sequence 107, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND
; COMPOSITIONS FOR INHIBITION
; OF MEMBRANE
; FUSION-ASSOCIATED EVENTS, ENCOMPASSING HIV

US-08-470-896-107
 ; Sequence 107, Application US/08470896
 ; Patent No. 6479055
 ;
 ; GENERAL INFORMATION:
 ;
 ; APPLICANT: Bolognesi, Dani P.
 ; APPLICANT: Matthews, Thomas J.
 ; APPLICANT: Wild, Carl T.
 ; APPLICANT: Barney, Shawn O.
 ; APPLICANT: Lambert, Dennis M.
 ; APPLICANT: Patteway, Stephen R.
 ; APPLICANT: Languois, Alphonse J.
 ;
 ; TITLE OF INVENTION: METHODS AND
 ; COMPOSITIONS FOR INHIBITION
 ; OF MEMBRANE
 ; FUSION-ASSOCIATED EVENTS, ENCOMPASSING HIV

1 PATENT NO. 6479035
 2
 3 GENERAL INFORMATION:
 4
 5 APPLICANT: Bolognesi, Dani P.
 6 APPLICANT: Matthews, Thomas J.
 7 APPLICANT: Wild, Carl T.
 8 APPLICANT: Barney, Shawn O.
 9 APPLICANT: Lambert, Dennis M.
 10 APPLICANT: Petteway, Stephen R.
 11 APPLICANT: Langlois, Alphonse J.
 12
 13 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
 14 OF MEMBRANE FUSION-ASSOCIATED EVENTS, ENCOMPASSING HIV

APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE FUSION-ASSOCIATED EVENTS, ENCOMPASSING HIV

APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE FUSION-ASSOCIATED EVENTS.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE FUSION-ASSOCIATED EVENTS.

APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway, Stephen R.
APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
OF MEMBRANE FUSION-ASSOCIATED EVENTS, ENCOMPASSING HIV

APPLICANT: Langlois, Alphonse J.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV

TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV

Query Match

Best Local Similarity 45.9%; Pred. No. 1.6e-05;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:37:23 ; Search time 54 Seconds
(without alignments)

476.145 Million cell updates/sec

Title: US-10-069-883-1_COPY_448_538

Perfect score: 457

Sequence: 1 ILPLGLPLAIIILLRGPCC.....PPETISAAQLLRNSAGSS 91

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|-------------|
| 1 | 457 | 100.0 | 538 | 2 | AA08622 | Human sec |
| 2 | 457 | 100.0 | 538 | 3 | AA067313 | Human sec |
| 3 | 457 | 100.0 | 538 | 4 | AA075138 | HERV-W en |
| 4 | 457 | 100.0 | 538 | 4 | AA067652 | Amino aci |
| 5 | 457 | 100.0 | 538 | 5 | AA014540 | Human syn |
| 6 | 457 | 100.0 | 538 | 7 | ADC38777 | Human sec |
| 7 | 457 | 100.0 | 540 | 3 | ABP40980 | Human ret |
| 8 | 457 | 100.0 | 544 | 7 | AE00243 | Novel pro |
| 9 | 457 | 100.0 | 685 | 3 | ABP40988 | Human ret |
| 10 | 457 | 100.0 | 846 | 3 | ABP41050 | HERV-7g e |
| 11 | 452 | 98.9 | 538 | 2 | AA093145 | Human end |
| 12 | 387 | 84.7 | 541 | 3 | AA081195 | Amino aci |
| 13 | 386 | 84.5 | 542 | 5 | AA025054 | Human EMB |
| 14 | 377 | 82.5 | 633 | 4 | ABG28306 | Novel hum |
| 15 | 373 | 81.6 | 368 | 4 | ABG28311 | Novel hum |
| 16 | 370 | 81.0 | 531 | 5 | ABP63036 | Human pol |
| 17 | 353 | 77.2 | 162 | 2 | AA071069 | Multiple |
| 18 | 353 | 77.2 | 162 | 2 | AA099554 | Protein e |
| 19 | 353 | 77.2 | 542 | 3 | AA019069 | Amino aci |
| 20 | 344 | 75.3 | 162 | 7 | ADB84404 | MSRV-1 as |
| 21 | 301.5 | 65.0 | 87 | 4 | ABG15631 | Novel hum |
| 22 | 301.5 | 65.0 | 87 | 4 | ABG27347 | Novel hum |
| 23 | 298 | 65.2 | 144 | 4 | ABG08534 | Novel hum |
| 24 | 293 | 64.1 | 103 | 6 | ABR41810 | Human DIT |
| 25 | 287 | 62.8 | 77 | 2 | AA071067 | Multiple |

| | | | | | | |
|----|-------|------|------|---|----------|-----------|
| 26 | 287 | 62.8 | 77 | 2 | AA099552 | Protein e |
| 27 | 287 | 62.8 | 77 | 7 | ADB84398 | MSRV-1 as |
| 28 | 267 | 58.4 | 52 | 2 | AA099346 | Human end |
| 29 | 259.5 | 56.8 | 361 | 4 | ABG05318 | Novel hum |
| 30 | 257 | 56.2 | 58 | 4 | ABG24204 | Novel hum |
| 31 | 253 | 55.4 | 641 | 4 | ABG09787 | Novel hum |
| 32 | 253 | 55.4 | 960 | 7 | ADC31280 | Human nov |
| 33 | 253 | 55.4 | 1067 | 4 | ABG08049 | Novel hum |
| 34 | 253 | 55.4 | 1288 | 4 | ABG04434 | Novel hum |
| 35 | 253 | 55.4 | 1288 | 4 | ABG12277 | Novel hum |
| 36 | 249 | 54.5 | 426 | 4 | ABG28313 | Novel hum |
| 37 | 247.5 | 54.2 | 231 | 4 | ABG12209 | Novel hum |
| 38 | 243 | 53.2 | 426 | 4 | ABG08531 | Novel hum |
| 39 | 238.5 | 52.2 | 1285 | 4 | ABG05808 | Novel hum |
| 40 | 235 | 51.4 | 475 | 4 | ABG28310 | Novel hum |
| 41 | 235 | 51.4 | 485 | 4 | ABG13126 | Novel hum |
| 42 | 229 | 50.1 | 641 | 4 | ABG08043 | Novel hum |
| 43 | 225 | 49.2 | 115 | 4 | ABG12822 | Novel hum |
| 44 | 224 | 49.0 | 143 | 4 | ABG03413 | Novel hum |
| 45 | 224 | 49.0 | 340 | 4 | ABG04438 | Novel hum |
| 46 | 224 | 49.0 | 559 | 4 | ABG03408 | Novel hum |
| 47 | 220 | 48.1 | 110 | 4 | ABG08040 | Novel hum |
| 48 | 220 | 48.1 | 312 | 4 | ABG14848 | Novel hum |
| 49 | 220 | 48.1 | 317 | 4 | ABG26803 | Novel hum |
| 50 | 220 | 48.1 | 317 | 4 | ABG08036 | Novel hum |
| 51 | 220 | 48.1 | 682 | 4 | ABG26804 | Novel hum |
| 52 | 217 | 47.5 | 123 | 4 | ABG04433 | Novel hum |
| 53 | 217 | 47.5 | 166 | 4 | ABG20016 | Novel hum |
| 54 | 216 | 47.3 | 158 | 4 | ABG13182 | Novel hum |
| 55 | 215.5 | 47.2 | 248 | 4 | ABG05807 | Novel hum |
| 56 | 215 | 47.0 | 571 | 4 | ABG09513 | Novel hum |
| 57 | 214 | 46.8 | 284 | 4 | ABG27348 | Novel hum |
| 58 | 213 | 46.6 | 476 | 4 | ABG26802 | Novel hum |
| 59 | 212 | 46.4 | 259 | 4 | ABG09734 | Novel hum |
| 60 | 212 | 46.4 | 455 | 4 | ABG08528 | Novel hum |
| 61 | 212 | 46.4 | 455 | 4 | ABG09503 | Novel hum |
| 62 | 212 | 46.4 | 536 | 4 | ABG12830 | Novel hum |
| 63 | 212 | 46.4 | 604 | 4 | ABG08533 | Novel hum |
| 64 | 210 | 46.0 | 592 | 4 | ABG28113 | Novel hum |
| 65 | 206 | 45.1 | 713 | 4 | ABG14831 | Novel hum |
| 66 | 206 | 45.1 | 741 | 4 | ABG03406 | Novel hum |
| 67 | 206 | 45.1 | 741 | 4 | ABG12276 | Novel hum |
| 68 | 206 | 45.1 | 765 | 4 | ABG02298 | Novel hum |
| 69 | 204 | 44.6 | 63 | 4 | AA017796 | Peptide # |
| 70 | 204 | 44.6 | 63 | 4 | ABG36821 | Peptide # |
| 71 | 204 | 44.6 | 63 | 4 | AA030306 | Peptide # |
| 72 | 204 | 44.6 | 63 | 4 | ABG31600 | Peptide # |
| 73 | 204 | 44.6 | 63 | 4 | ABG22144 | Protein # |
| 74 | 204 | 44.6 | 63 | 4 | AA069965 | Human bon |
| 75 | 204 | 44.6 | 63 | 4 | AA057562 | Human bra |
| 76 | 204 | 44.6 | 63 | 4 | AA051666 | Human liv |
| 77 | 204 | 44.6 | 63 | 4 | AA05443 | Peptide # |
| 78 | 204 | 44.6 | 63 | 5 | ABG39598 | Human pep |
| 79 | 204 | 44.6 | 433 | 2 | AA036029 | Protein e |
| 80 | 204 | 44.6 | 433 | 2 | AA036029 | Protein e |
| 81 | 204 | 44.6 | 433 | 7 | ADB84376 | MSRV-1 as |
| 82 | 203 | 44.4 | 1415 | 4 | ABG14784 | Novel hum |
| 83 | 196 | 42.9 | 68 | 4 | AA019429 | Human dia |
| 84 | 194 | 42.5 | 591 | 4 | ABG09510 | Novel hum |
| 85 | 191 | 41.8 | 953 | 4 | ABG03420 | Novel hum |
| 86 | 189 | 41.4 | 493 | 2 | AA071068 | Multiple |
| 87 | 189 | 41.4 | 493 | 2 | AA099553 | Protein e |
| 88 | 189 | 41.4 | 493 | 7 | ADB84401 | MSRV-1 as |
| 89 | 188 | 41.1 | 716 | 4 | ABG10593 | Novel hum |
| 90 | 188 | 41.1 | 731 | 4 | ABG05333 | Novel hum |
| 91 | 188 | 41.1 | 875 | 4 | ABG06222 | Novel hum |
| 92 | 188 | 41.1 | 876 | 4 | ABG04569 | Novel hum |
| 93 | 185 | 40.5 | 59 | 4 | ABG16284 | Novel hum |
| 94 | 184 | 40.3 | 153 | 4 | ABG04978 | Novel hum |
| 95 | 183 | 40.0 | 1361 | 4 | ABG17354 | Novel hum |
| 96 | 183 | 40.0 | 1361 | 4 | ABG13631 | Novel hum |
| 97 | 179 | 39.2 | 244 | 4 | ABG17351 | Novel hum |
| 98 | 175 | 38.3 | 50 | 4 | AA015669 | Peptide # |

99 175 38.3 50 4 AAM28170 Peptide #
 100 175 38.3 50 4 ABB20079 Protein #

ALIGNMENTS

RESULT 1
 AAY08622
 ID AAY08622 standard; protein; 538 AA.

AC AAY08622;
 XX 10-AUG-1999 (first entry)
 XX Human secreted protein AJ172_2.
 DE
 XX Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
 KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
 KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
 KW haematopoiesis regulator; activin; inhibitor; chemotactic; chemokinetic;
 KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
 KW cadherin; tumour invasion suppressor; gene therapy; tissue growth.
 XX
 OS Homo sapiens.
 XX
 PN WO9926972-A1.
 XX
 XX 03-JUN-1999.
 XX
 XX 17-NOV-1998; 98WO-US024614.
 XX
 XX 21-NOV-1997; 97US-00976110.
 PR 18-MAY-1998; 98US-00080478.
 PR 20-OCT-1998; 98US-00175928.
 XX
 XX (GEMY) GENETICS INST INC.
 PA
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 PI
 XX WPI; 1998-357813/30.
 DR N-PSDB; AAX77526.
 XX

PT New polynucleotides encoding secreted proteins.

PS Claim 148; Page 101-103; 142pp; English.

CC This invention describes novel human secreted proteins encoded by
 CC polynucleotides isolated from human adult testes, adult brain, adult
 CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
 CC libraries. The products of the invention are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional
 CC activity, cytokine and cell proliferation/differentiation activity,
 CC immune stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, and tumour
 CC inhibition activity. The polynucleotides are also stated to be useful for
 CC gene therapy
 XX

SQ Sequence 538 AA;

Query Match 100.0%; Score 457; DB 2; Length 538;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKQSKTKIYRRPLDRPAS 60
 DB 448 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKQSKTKIYRRPLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
 DB 508 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 538

RESULT 2

AAY67313
 ID AAY67313 standard; protein; 538 AA.

XX AAY67313;
 AC AAY67313;
 XX 11-APR-2000 (first entry)
 XX Human secreted protein AJ172_2 amino acid sequence.
 DE
 XX Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
 KW placental pathology; metastasis inhibition; nutritional activity;
 KW immune stimulator; haematopoiesis regulator; tissue growth;
 KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115.
 XX
 OS Homo sapiens.
 XX
 PN WO9960020-A1.
 XX
 XX 25-NOV-1999.
 XX
 XX 17-MAY-1999; 99WO-US010915.
 XX
 XX 18-MAY-1998; 98US-00080478.
 PR 20-OCT-1998; 98US-00175928.
 XX
 XX (GEMY) GENETICS INST INC.
 PA
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Mi S, Treacy M;
 PI
 XX WPI; 2000-116311/10.
 DR N-PSDB; AAZ59468.
 XX

PT New polynucleotides encoding secreted cDNA libraries, used to develop
 products for the diagnosis and treatment of neoplastic disease.

PS Claim 15; Page 108-110; 149pp; English.

CC This is the human secreted protein AJ172_2 amino acid sequence, the
 CC polynucleotide encoding AJ172_2 was obtained from a human adult testes
 CC cDNA library. The invention relates to secreted human and murine
 CC proteins. The polynucleotides and proteins are predicted to have
 CC biological activities which would make them suitable for treating,
 CC preventing or ameliorating medical conditions in humans and animals.
 CC Detection of the levels of the proteins can be used for the diagnosis of
 CC e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate
 CC the expression or function of the proteins may be used for treating a
 CC neoplastic disease and inhibiting metastasis. Other suggested activities
 CC include nutritional activity (e.g. in feeds), cytokine and cell
 CC proliferation/differentiation activity, immune stimulating (e.g. as
 CC vaccines) or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The
 CC polynucleotide sequences are also stated to be useful for gene therapy
 XX

SQ Sequence 538 AA;

Query Match 100.0%; Score 457; DB 3; Length 538;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKQSKTKIYRRPLDRPAS 60
 DB 448 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLMQEPKQSKTKIYRRPLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 91
 DB 508 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 538

RESULT 3

AAB75138
 ID AAB75138 standard; protein; 538 AA.

AC AAB75138;

DT 08-AUG-2001 (first entry)

DE HERV-W envelope protein G.

XX Human endogenous retrovirus; HERV-W; chromosome 7; env protein;
 XX envelope protein; multiple sclerosis-related superantigen; vaccine;
 KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
 KW antisense-therapy; autoimmune disorder.

XX Human endogenous retrovirus.

XX Key Location/Qualifiers

FT Region 271..317

FT /label= SU

FT /note= "surface protein"

FT Cleavage-site 317..318

FT /note= "SU-TM putative cleavage site"

FT Region 318..538

FT /label= TM

FT /note= "transmembrane subunit"

XX WO200131021-A1.

XX 03-MAY-2001.

XX 30-OCT-2000; 2000WO-EP010659.

XX 28-OCT-1999; 99EP-00402690.

XX (UYGE-) UNIV GENEVE.

XX Conrad B, Mach B;

XX WPI; 2001-316336/33.

XX N-PSDB; AAH20070.

XX New human retrovirus HERV-W ENV proteins/peptides having superantigen
 PT activity useful for diagnosing and treating multiple sclerosis.
 XX Claim 1; Fig 7; 94pp; English.

On the basis of the PBS t-RNA motif used for the classification of human
 endogenous retrovirus (HERVs) the full length endogenous provirus which
 was been located on the long arm of human chromosome 7 (7q21-22) has been
 designated HERV-W. The present invention describes proteins or peptides
 (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
 of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
 have neuroprotective activity, and can be used in: vaccines; antisense-
 therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
 useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
 disorders. (I) are also useful for identifying substances (and optionally
 recovering), capable of binding to a retroviral superantigen associated
 with MS, substances capable of blocking SAG activity and substances
 capable of blocking transcription or translation of HERV-W retroviral
 superantigen. A protein or peptide derived from (I), modified to be
 devoid of SAG activity and being capable of generating an immune response
 against HERV-W retroviral SAG is useful in therapy. Nucleic acid
 molecules encoding (I) are useful as vaccines against MS. Substances
 capable of blocking SAG activity, capable of binding to a retroviral
 superantigen associated with MS, or capable of blocking transcription or
 translation of HERV-W retroviral superantigen for use in treating or

CC preventing MS, obtained using (I) are useful for the treatment and
 CC prevention of MS. (I) and nucleic acids encoding them are useful for
 CC diagnosing autoimmune disease. The present sequence represents the
 CC specifically claimed envelope protein of HERV-W designated G

XX Sequence 538 AA;

Query Match 100.0%; Score 457; DB 4; Length 538;

Best Local Similarity 100.0%; Pred. No. 1.2e-45;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPQMOSKTKIYRRPLDRPAS 60

DB 448 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPQMOSKTKIYRRPLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 91

DB 508 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 538

RESULT 4

AAB67652

ID AAB67652 standard; protein; 538 AA.

XX AAB67652;

DT 29-MAY-2001 (first entry)

XX Amino acid sequence of a human endogenous retrovirus envelope protein.
 KW Envelope protein; HERV; syncytia formation; placental development;
 KW syncytia; cancer; cell adhesion.
 XX Human endogenous retrovirus.
 XX WO200116171-A1.
 XX 08-MAR-2001.
 XX 01-SEP-2000; 2000WO-FR002429.
 XX 01-SEP-1999; 99FR-00011141.
 XX 15-SEP-1999; 99FR-00011793.
 XX (INMR) BIO MERIEUX.
 XX (INRM) INST NAT SANTE & RECH MEDICALE.
 XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
 WPI; 2001-226676/23.
 XX N-PSDB; AAF55830.

XX Detecting expression of human endogenous retrovirus envelope protein in
 cells of a tissue or culture, from its ability to induce syncytia.
 XX Claim 1; Page 41-44; 57pp; French.
 The present sequence represents a human endogenous retrovirus envelope
 protein. The specification describes a method for detecting expression of
 an envelope protein from a human endogenous retrovirus (HERV), in cells,
 of a tissue or culture. The method comprises detecting syncytia formation
 due to the fusogenic properties of the envelope protein. Envelope
 polypeptides and polynucleotides are used to produce therapeutic or
 prophylactic compositions, particularly for treatment of cancer, to
 correct defects in placental development (or other natural formation of
 other types of syncytia), and to promote adhesion of cells in grafts or
 cellular repair processes. Expression of sequences antisense to the
 polynucleotide are used to prevent formation of syncytia

XX Sequence 538 AA;

Query Match 100.0%; Score 457; DB 4; Length 538;

Best Local Similarity 100.0%; Pred. No. 1.2e-45;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPLGLPAAIILLFGPCIFNLVNFVSSRIEAVKLOMEPKMQSKTKIYRRLDRPAS 60
 Db 448 ILPLGLPAAIILLFGPCIFNLVNFVSSRIEAVKLOMEPKMQSKTKIYRRLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
 Db 508 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

RESULT 5
 AAE14540
 ID AAE14540 standard; protein; 538 AA.
 AC AAE14540;
 XX 07-MAY-2002 (first entry)
 DT Human syncytin protein.
 XX Human; syncytin; preclampsia; gestational trophoblast disorder;
 KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;
 KW envelope gene; human endogenous defective retrovirus; HERV-W.
 XX Homo sapiens.
 OS
 XX WO200204678-A2.
 PN 17-JAN-2002.
 XX 09-JUL-2001; 2001WO-US021719.
 PF 07-JUL-2000; 2000US-0216657P.
 PR (GEMY) GENETICS INST INC.
 PA Keith JC, Mccoy JM, Mi S;
 PI WPI; 2002-171727/22.
 DR N-PSDB; AAD24195.
 XX Identifying a compound for treating a subject with or at risk of
 PT developing preclampsia, comprises determining whether the expression or
 PT activity of syncytin in the cell is modulated in the presence of a test
 PT compound.
 XX Disclosure; Page 42-43; 43pp; English.
 XX The invention relates to identifying compounds which are modulators of
 CC syncytin expression. The syncytin modulators are useful in diagnosis and
 CC treatment of preclampsia and gestational trophoblast disorders (e.g.
 CC choriocarcinoma, hydatiform mole, placental site tumour and missed/
 CC incomplete abortion). Syncytin is a human gene derived from the envelope
 CC gene of human endogenous defective retrovirus, HERV-W. The present
 CC invention is based partly on the discovery that syncytin expression is
 CC dramatically reduced in preclampsia, and is also mis-localised to the
 CC apical syncytiotrophoblast membrane. The present sequence is human
 CC syncytin protein
 XX Sequence 538 AA;
 SQ

Query Match 100.0%; Score 457; DB 5; Length 538;
 Best Local Similarity 100.0%; Pred. No. 1.2e-45;
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPLGLPAAIILLFGPCIFNLVNFVSSRIEAVKLOMEPKMQSKTKIYRRLDRPAS 60
 Db 448 ILPLGLPAAIILLFGPCIFNLVNFVSSRIEAVKLOMEPKMQSKTKIYRRLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
 Db 508 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

RESULT 6
 ADC38777
 ID ADC38777 standard; protein; 538 AA.
 XX ADC38777;
 AC ADC38777;
 XX 18-DEC-2003 (first entry)
 DT Human secreted protein #63.
 XX
 DE immune disorder; severe combined immunodeficiency; SCID;
 KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
 KW rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
 KW lymphoid cell deficiency; osteoporosis; osteoarthritis;
 KW peripheral nervous system disease; peripheral neuropathy;
 KW Alzheimer's disease; Parkinson's disease; coagulation disorder;
 KW inflammatory disease; systemic inflammatory response syndrome; SIRS;
 KW ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
 KW hypersensitivity; regeneration; neural cell proliferation; fertility;
 KW tumour; chemokine; human; secreted protein.
 XX Homo sapiens.
 OS
 XX US2002193567-A1.
 PN 19-DEC-2002.
 XX 02-APR-2002; 2002US-00114893.
 PF 11-AUG-1995; 95US-00514014.
 PR 05-APR-1996; 96US-00628364.
 PR 19-APR-1996; 96US-00635311.
 PR 07-JUN-1996; 96US-00659224.
 PR 17-JUN-1996; 96US-00664596.
 PR 09-JUL-1996; 96US-00677231.
 PR 26-JUL-1996; 96US-00686878.
 PR 23-AUG-1996; 96US-00701819.
 PR 27-SEP-1996; 96US-00721488.
 PR 27-SEP-1996; 96US-00721798.
 PR 27-SEP-1996; 96US-00721923.
 PR 25-OCT-1996; 96US-00721926.
 PR 30-OCT-1996; 96US-00738367.
 PR 13-JAN-1997; 97US-00739775.
 PR 10-APR-1997; 97US-00833823.
 PR 02-JUN-1997; 97US-00867677.
 PR 05-SEP-1997; 97US-00924838.
 PR 06-OCT-1999; 99US-00413232.
 XX (GEMY) GENETICS INST INC.
 XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Buckett M;
 PI Kelleher K;
 XX WPI; 2003-657236/62.
 DR N-PSDB; ADC38776.
 XX Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
 PT BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
 PT useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
 XX Disclosure; SEQ ID NO 135; 412pp; English.
 XX The invention relates to a protein comprising fully defined AZ302 1
 CC protein or BD127 1 6 protein. The polynucleotides are useful for
 CC expressing recombinant proteins for analysis and are also useful as
 CC chromosome markers or tags to identify chromosomes or to map related gene
 CC positions. The proteins are useful as amino acid supplement, carbon
 CC source, nitrogen source and carbohydrate source. The proteins are useful
 CC for treating various immune deficiencies and disorders (e.g. severe

CC combined immunodeficiency (SCID), autoimmune disorders (e.g. multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic reactions (e.g. asthma, myeloid or lymphoid cell deficiencies, osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g. peripheral neuropathy, Alzheimer's disease, Parkinson's disease), coagulation disorders, inflammatory diseases (e.g. systemic inflammatory response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease), anaphylaxis and hypersensitivity. Proteins are also useful for inducing tumour immunity, for inducing bone, cartilage, tendon, ligament and/or nerve growth or regeneration, for proliferating neural cells and for regenerating nerve and brain tissue, for inducing fertility and for inhibiting tumour growth. Proteins are also useful as chemokine for mammalian cells (e.g. monocytes, fibroblasts, neutrophils), and also useful as inhibitors of receptor/ligand interactions. The present sequence represents the amino acid sequence of a human secreted protein.

XX SQ Sequence 538 AA;

Query Match 100.0%; Score 457; DB 7; Length 538;
Best Local Similarity 100.0%; Pred. No. 1.2e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLOMEPRKQSKTKIYRPLDRPAS 60
DB 448 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLOMEPRKQSKTKIYRPLDRPAS 507

QY 61 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
DB 508 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 538

RESULT 7

ABP40980
ID ABP40980 standard; protein; 540 AA.

XX AC ABP40980;

DT 01-AUG-2002 (first entry)

DE Human retroviral HERV-7q env peptide #3.

KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
multiple sclerosis.

OS Human endogenous retrovirus.

PN W09567395-A1.

PD 29-DEC-1999.

PF 23-JUN-1999; 99WO-FR001513.

PR 23-JUN-1998; 98FR-00007920.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Alliel PM, Perin J, Rieger F;

DR WPI; 2000-160587/14.

PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used for diagnosis, treatment and prevention of autoimmune and neurological diseases.

PS Claim 22; Fig 4; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human endogenous retrovirus, HERV-7q, which is located on chromosome 7q. CC Regulatory elements associated with HERV-7q may alter expression of other CC genes (even remote genes) on the same chromosome, inducing immunological CC and/or neurological changes (which may be pathological or protective/ curative). HERV-7q peptides can be used to improve efficiency of the CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding

CC sequences can be used in immunogenic or vaccinating compositions, for protection against autoimmune diseases, particularly multiple sclerosis. The peptides may also be used (by sequence comparison) to detect/identify CC endogenous retroviruses that are abnormally expressed in cancer, CC neuropathologies or other autoimmune diseases. The present sequence was CC used to illustrate the invention

XX SQ Sequence 540 AA;

Query Match 100.0%; Score 457; DB 3; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.3e-45;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLOMEPRKQSKTKIYRPLDRPAS 60
DB 450 ILPLGLPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLOMEPRKQSKTKIYRPLDRPAS 509

QY 61 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
DB 510 PRSDVNDIKGTPPEISAAQPLLRPNAGSS 540

RESULT 8
ADE09243
ID ADE09243 standard; protein; 544 AA.

XX AC ADE09243;

XX DT 29-JAN-2004 (first entry)

DE Novel protein-related contig polypeptide sequence #309.

XX novel gene; novel protein; tissue marker; molecular weight marker;
XX chromosome marker; genetic disorder; contig.

OS Unidentified.

XX WO2003054152-A2.

XX PD 03-JUL-2003.

PF 10-DEC-2002; 2002WO-US039555.

PR 10-DEC-2001; 2001US-0339739P.

PR 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.

PR 12-APR-2002; 2002US-0372381P.

PR 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.

PR 24-APR-2002; 2002US-0376045P.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Dmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for analysis, characterization or therapeutic use, or as markers for tissues in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2787; 1177pp; English.

CC The invention comprises the amino acid and coding sequences of novel CC proteins. The DNA and protein sequences of the invention are useful as: CC markers for tissues in which the corresponding protein is preferentially CC expressed; as molecular weight markers on gels; as chromosome markers or CC tags; to identify chromosomes or to map related gene positions; and to CC compare with endogenous DNA sequences in patients to identify potential CC genetic disorders. The present amino acid sequence was used in the

CC exemplification of the invention.

SQ Sequence 544 AA;

Query Match 100.0%; Score 457; DB 7; Length 544;

Best Local Similarity 100.0%; Pred. No. 1.3e-45; Indels 0; Gaps 0; Mismatches 0; Conservative 0;

Qy 1 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 60

Db 454 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 513

Qy 61 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 91

Db 514 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 544

RESULT 9

ID ABP40988

XX ABP40988 standard; protein; 685 AA.

AC ABP40988;

XX 01-AUG-2002 (first entry)

XX Human retroviral env ORF protein.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX Multiple sclerosis.

XX Human endogenous retrovirus.

XX WO9967395-A1.

XX 29-DEC-1999.

XX 23-JUN-1999; 99WO-FR001513.

XX 23-JUN-1998; 98FR-00007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used

XX for diagnosis, treatment and prevention of autoimmune and neurological

XX diseases.

XX Claim 22; Page 152-154; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human

XX endogenous retrovirus, HERV-7q, which is located on chromosome 7q.

XX Regulatory elements associated with HERV-7q may alter expression of other

XX genes (even remote genes) on the same chromosome, inducing immunological

XX and/or neurological changes (which may be pathological or protective/

XX curative). HERV-7q peptides can be used to improve efficiency of the

XX immune response, e.g. in immunotherapy. HERV-7q peptides and their coding

XX sequences can be used in immunogenic or vaccinating compositions, for

XX protection against autoimmune diseases, particularly multiple sclerosis.

XX The peptides may also be used (by sequence comparison) to detect/identify

XX endogenous retroviruses that are abnormally expressed in cancer,

XX neuropathologies or other autoimmune diseases. The present sequence was

XX used to illustrate the invention

SQ Sequence 685 AA;

Query Match

Best Local Similarity 100.0%; Score 457; DB 3; Length 685;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 60

Db 578 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 637

Qy 61 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 91

Db 638 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 668

RESULT 10

ID ABP41050

XX ABP41050 standard; protein; 846 AA.

AC ABP41050;

XX 01-AUG-2002 (first entry)

XX HERV-7q env protein #1.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX Multiple sclerosis.

XX Human endogenous retrovirus.

XX WO9967395-A1.

XX 29-DEC-1999.

XX 23-JUN-1999; 99WO-FR001513.

XX 23-JUN-1998; 98FR-00007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used

XX for diagnosis, treatment and prevention of autoimmune and neurological

XX diseases.

XX Example 4; Fig 19; 225pp; French.

The present invention relates to new nucleic acid sequences of human endogenous retrovirus, HERV-7q, which is located on chromosome 7q. Regulatory elements associated with HERV-7q may alter expression of other genes (even remote genes) on the same chromosome, inducing immunological and/or neurological changes (which may be pathological or protective/curative). HERV-7q peptides can be used to improve efficiency of the immune response, e.g. in immunotherapy. HERV-7q peptides and their coding sequences can be used in immunogenic or vaccinating compositions, for protection against autoimmune diseases, particularly multiple sclerosis. The peptides may also be used (by sequence comparison) to detect/identify endogenous retroviruses that are abnormally expressed in cancer, neuropathologies or other autoimmune diseases. The present sequence was used to illustrate the invention

SQ Sequence 846 AA;

Query Match

Best Local Similarity 100.0%; Score 457; DB 3; Length 846;

Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 60

Db 738 ILPLGFLAAIILLLLFGPCIFNLLVNFVSSRIEAVKLQMEPKMQSKTKIYRRPLDRPAS 797

Qy 61 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 91

Db 798 PRSDVNDIKGTPPEISAAQPLLRPSAGSS 828

RESULT 11

AAW99345
 ID AAW99345 standard; protein; 538 AA.
 XX
 AC AAW99345;
 XX
 DT 21-MAY-1999 (first entry)
 XX
 DE Human endogenous retrovirus W ORF1.
 XX
 KW Clone; human endogenous retrovirus; genome; autoimmune disease;
 KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker.
 XX
 OS Human endogenous retrovirus.
 XX
 PN WO9902696-A1.
 XX
 PD 21-JAN-1999.
 XX
 PF 06-JUL-1998; 98WO-FR001442.
 XX
 PR 07-JUL-1997; 97FR-00008815.
 XX
 PA (INMR) BIO MERIEUX.
 XX
 PI Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;
 XX
 DR WPI; 1999-120897/10.
 XX
 PT New nucleic acid sequences from human endogenous retrovirus-W - expressed
 PT exclusively in placenta and useful in diagnosis and therapy of autoimmune
 PT disease, and abnormal or failed pregnancy.
 XX
 PS Example 8; Fig 6; 106pp; French.
 XX
 CC This sequence represents the open reading frame 1 encoding the env
 CC protein from the human endogenous retrovirus (HERV) W (AAW25665). Nucleic
 CC acids, their fragments or peptides encoded by them derived from the HERV-
 CC W genome are markers of autoimmune disease (e.g. multiple sclerosis,
 CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
 CC dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility
 XX
 SQ Sequence 538 AA;
 XX
 Query Match 98.9%; Score 452; DB 2; Length 538;
 Best Local Similarity 98.9%; Pred. No. 4, 9e-45;
 Matches 90; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKQLQMEPKQSKTKIYRRLDRPAS 60
 DB 448 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKQLQMEPKQSKTKIYRRLDRPAS 507
 QY 61 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
 DB 508 PRSDVNDIKGTPPEISAAQPLLRNSAGSS 538
 RESULT 12
 ID AAB08195
 XX
 AC AAB08195;
 XX
 DT 06-AUG-2003 (revised)
 DT 04-DEC-2000 (first entry)
 XX
 DE Amino acid sequence of the MSRV-1 3' env and LTR regions.
 XX
 KW MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus.
 XX

OS Multiple sclerosis associated retrovirus.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT Modified-site /note= "signal peptide"
 FT Modified-site 58..60
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site 92..94
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site 213..215
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site 233..235
 FT Modified-site /note= "N-linked glycosylation site"
 FT Modified-site 280..282
 FT Modified-site /note= "N-linked glycosylation site"
 FT Cleavage-site 289..290
 FT Cleavage-site /note= "putative cleavage site"
 FT Cleavage-site 315..316
 FT Modified-site /note= "putative cleavage site"
 FT Modified-site 408..410
 FT Modified-site /note= "N-linked glycosylation site"
 XX
 PN WO200047745-A1.
 XX
 PD 17-AUG-2000.
 XX
 PF 15-FEB-2000; 2000WO-IB000159.
 XX
 PR 15-FEB-1999; 99EP-00420041.
 XX
 PA (INMR) BIO MERIEUX.
 XX
 PI Paranhos-Baccala G, Perron H, Komurian-Pradel F;
 XX
 DR WPI; 2000-506097/45.
 DR N-PSDB; AAA63826.
 XX
 PT Nucleotide fragment of LTR-RUS region from Multiple Sclerosis retrovirus
 PT (MSRV) used to detect the presence of MSRV-1 retrovirus in a biological
 PT sample.
 XX
 PS Disclosure; Fig 2; 23pp; English.
 XX
 CC The present sequence is encoded by a nucleotide sequence corresponding
 CC to the 3' env region and long terminal repeat sequences from clone CL6 of
 CC Multiple Sclerosis retrovirus (MSRV-1). The specification describes a
 CC long terminal repeat (LTR)-RUS region which encodes the expression of a
 CC MSRV-1 protein. This is unusual for LTRs, in particular in the RUS
 CC region. Probes and antibodies to the MSRV-1 retrovirus protein and
 CC encoding polynucleotide sequences are used to detect the presence of MSRV
 CC -1 retrovirus in a biological sample. (Updated on 06-AUG-2003 to correct
 CC OS field.)
 XX
 SQ Sequence 541 AA;
 XX
 Query Match 84.7%; Score 387; DB 3; Length 541;
 Best Local Similarity 83.2%; Pred. No. 2, 9e-37;
 Matches 79; Conservative 6; Mismatches 6; Indels 4; Gaps 1;
 QY 1 ILPFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVK---LQMEPKQSKTKIYRRLD 56
 DB 447 VLPLFLGPLAAIILLLLFGPCIFNLLVNFVSSRIEAVKQLQMEPKQSKTKIHRGPLD 506
 QY 57 RPASPSDVNDIKGTPPEISAAQPLLRNSAGSS 91
 DB 507 RPASPSDVNDIKGTPPEISAAQPLLRNSAGSS 541
 RESULT 13
 ID AAE25054
 ID AAE25054 standard; protein; 542 AA.
 XX
 AC AAE25054;
 XX

XX 30-OCT-2002 (first entry)
XX Human EMBRY-2 protein.
XX
XX Human; embryogenesis associated protein; AIDS; reproductive disorder;
XX infertility; endometriosis; endometrial tumour; inflammatory disorder;
XX autoimmune disorder; acquired immune deficiency syndrome; transgenic;
XX ovarian tumour; contact dermatitis; placenta disorder; preeclampsia;
XX EMBRY-2; allergy; gene therapy.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..20
XX /label= Signal_peptide
XX Protein 21..542
XX /note= "Mature EMBRY-2 protein"
XX Domain 263..484
XX /note= "ENV polyprotein (coat polyprotein) domain"
XX Domain 455..473
XX /note= "transmembrane domain"
XX
XX W0200248362-A2.
XX
XX 20-JUN-2002.
XX
XX 14-NOV-2001; 2001WO-US043956.
XX
XX 15-NOV-2000; 2000US-0249407P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Ramkumar J, Arvizu C;
XX WPI; 2002-537629/57.
XX N-PSDB; AAD41225.
XX
XX New polypeptides of human embryogenesis associated proteins for screening
XX modulators useful for treating or preventing disorders e.g.
XX endometriosis, infertility, allergy, preeclampsia.
XX
XX Claim 57; Page 94-95; 97pp; English.
XX
XX The invention relates to human embryogenesis associated proteins (EMBR)
XX and nucleic acid molecules encoding such proteins. EMBRY sequences are
XX useful for screening modulators useful for treating or preventing
XX disorders associated with abnormal expression of EMBRY. The disorders
XX treated include reproductive disorders such as infertility,
XX endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory
XX disorder such as acquired immune deficiency syndrome (AIDS), allergies,
XX contact dermatitis; disorders of the placenta such as preeclampsia,
XX abruptio placentae etc. Sequences of the invention are also useful for
XX analysing a proteome of a tissue or a cell type. EMBRY proteins are
XX useful as immunogens for preparing antibodies. Polynucleotides of the
XX invention are useful for creating knockin humanised animals or transgenic
XX animals to model human diseases. They are also used in gene therapy. The
XX present sequence is human EMBRY-2 protein
XX
XX Sequence 542 AA;
XX
XX Query Match 84.5%; Score 386; DB 5; Length 542;
XX Best Local Similarity 83.2%; Pred. No. 3.8e-37;
XX Matches 79; Conservative 5; Mismatches 7; Indels 4; Gaps 1;
XX
XX 1 ILFFGLPAAIILLFGPCIFNLVNFVSSRIEAVK----LQMEPKMQSKTIYRRPLD 56
XX 448 VLFFGLPAAIILLFGPCIFNLVNFVSSRIEAVKQLQWLQMEPKMQSKTIYHGPLD 507
XX
XX 57 RPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 91
XX 508 RPASPCSDVNDIEGTPPEISAAQPLLRNSAGSS 542

RESULT 14
ABG28306
ID ABG28306 standard; protein; 633 AA.
XX
XX AC ABG28306;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #28297.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS92493.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 58665; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 633 AA;
XX
XX Query Match 82.5%; Score 377; DB 4; Length 633;
XX Best Local Similarity 83.2%; Pred. No. 5.4e-36;
XX Matches 79; Conservative 3; Mismatches 9; Indels 4; Gaps 1;
XX
XX 1 ILFFGLPAAIILLFGPCIFNLVNFVSSRIEAVK----LQMEPKMQSKTIYRRPLD 56
XX 539 ILFFGLPAAIILLFGPCIFNLVNFVSSRIEAVKQLQWLQMEPKMQSKTIYHGPLD 598
XX
XX 57 RPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 91

Db 599 QPASPCSDVNDIKGTPPEEISIAQHLCPNSAGSS 633

RESULT 15
ABG28311
ID ABG28311 standard; protein; 368 AA.

XX AC ABG28311;
XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #28302.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS92498.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 58670; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 368 AA;

Query Match 81.6%; Score 373; DB 4; Length 368;
Best Local Similarity 82.1%; Pred. No. 8.3e-36;
Matches 78; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 1 ILPFLGPLAAIILLFLFGPCIFNLVNFVSSRIEAVK----LQMEPKMQSKTKIYRRPLD 56
DB 258 ILPFLGPLAAIILLFLFGPCIFNLVNFVSSRIEAVKILQILQMEPKMQSKTKIYRRPLD 317

QY 57 RPASPRSDVNDIKGTPPEEISIAAQPILLRPNAGSS 91
DB 318 QPASPCSDVNDIEGTPPEEISNAQPLLCPNXAGSS 352

RESULT 16

ABP63036

XX ID ABP63036 standard; protein; 531 AA.

XX AC ABP63036;

XX DT 14-OCT-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 473.

XX KW Human; vulnary; dermatological; neuroprotective; nootropic; cancer;
XX KW antiparkinsonian; immunostimulant; cytostatic; immunosuppressive;
XX KW antidiabetic; antiallergic; gene therapy; wound healing; tissue repair;
XX KW burn; central nervous system disorder; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease; immune disorder;
XX KW autoimmune disorder; multiple sclerosis; diabetes; allergy.

XX OS Homo sapiens.

XX PN WO200218424-A2.

XX PD 07-MAR-2002.

XX PF 31-AUG-2001; 2001WO-US027093.

XX PR 01-SEP-2000; 2000US-00654935.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
XX PI Zhao QA, Wang D, Liu C, Drmanac RT, Wehrman T;

XX DR WPI; 2002-583321/62.

XX DR N-PSDB; ABQ93515.

XX PT New polynucleotide and polypeptides, useful for treatment and diagnosis
XX PT of Alzheimer's, Parkinson's, Huntington's, amyotrophic lateral
XX PT sclerosis, immune deficiencies, cancer, autoimmune disorders, multiple
XX PT sclerosis, diabetes and allergies.

XX PS Claim 20; SEQ ID NO 473; 284pp + Sequence Listing; English.

XX CC The invention relates to an isolated polynucleotide (I) comprising one of
XX CC 245 sequences (ABQ93288-ABQ93532). Treating a condition comprising
XX CC administering to a mammalian subject a composition comprising the protein
XX CC (II) encoded by (I) (ABP62809-ABP63053) or an antibody (III) to (II).
XX CC (I), (II) and (III) are useful for diagnostic evaluation of disorders.
XX CC (I) is useful for gene therapy of diseases and (II) can be used for
XX CC therapeutic treatment. Diseases that may be treated include wound healing
XX CC and tissue repair, burns, central nervous system disorders (e.g.
XX CC Alzheimer's, Parkinson's, Huntington's and amyotrophic lateral
XX CC sclerosis), immune deficiencies, cancer, autoimmune disorders, multiple
XX CC sclerosis, diabetes and allergies. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 531 AA;

Query Match 81.0%; Score 370; DB 5; Length 531;
Best Local Similarity 81.9%; Pred. No. 3e-35;
Matches 77; Conservative 4; Mismatches 9; Indels 4; Gaps 1;

QY 1 ILPFLGPLAAIILLFLFGPCIFNLVNFVSSRIEAVK----LQMEPKMQSKTKIYRRPLD 56
DB 366 VLFPFLGPLAAIILLFLFGPCIFNLVNFVSSRIEAVKLQWLQMEPKMQSKTKIYRRPLD 425

QY 57 RPASPRSDVNDIKGTPPEEISIAAQPILLRPNAGSS 90

DE Amino acid sequence of an envelope (env) protein of MSRV-1.
 XX Envelope protein; MSRV-1; superantigen; autoimmune disease; Vbetals;
 KW Vbeta17; multiple sclerosis; vaccine.
 XX Human endogenous retrovirus.
 XX WO200057185-A1.
 XX 28-SEP-2000.
 XX 20-MAR-2000; 2000WO-FR000691.
 XX 19-MAR-1999; 99FR-00003622.
 XX 28-OCT-1999; 99FR-00013755.
 XX (INMR) BIO MERIEUX.
 XX Perron H, Lafont M;
 PI WPI; 2000-638279/61;
 DR N-PSDB; AAA96625.
 XX Detecting superantigen activity, useful for identifying agents for
 PT treatment or prevention of autoimmune disease, from expansion or loss of
 PT particular lymphocyte Vbeta determinants.
 XX Claim 28; Page 124-126; 134pp; French.
 XX The present sequence represents an envelope protein of MSRV-1. The
 CC superantigen protein expressed by the endogenous human retrovirus MSRV-1 has
 CC superantigen activity associated with autoimmune disease. The protein can
 CC be detected using the method of the invention. The specification
 CC describes a process for detecting activity of a superantigen in a
 CC biological sample. The process comprises identifying large scale
 CC expansion or loss of lymphocytes that carry at least one of the Vbeta16
 CC and/or Vbeta17 determinants. The method is used to screen for agents that
 CC inhibit the superantigen, especially those associated with MSRV-1 which
 CC is implicated in autoimmune disease, particularly multiple sclerosis.
 CC These agents are potentially useful for treatment or prevention (e.g. as
 CC vaccines) of autoimmune diseases
 XX Sequence 542 AA;
 SQ
 Query Match 77.2%; Score 353; DB 3; Length 542;
 Best Local Similarity 79.8%; Pred. No. 3.3e-33;
 Matches 75; Conservative 2; Mismatches 13; Indels 4; Gaps 1;
 QY 2 LPFLGPLAAIILLFPGPCIFNLLVNFVSSRIEAVK----LQMEPKMQSKTKIYRRPLDR 57
 DB 449 LPFLGPLAAIIFLLFPGPCIFNLFVSVSRIRIYVAVKQLQIVLQMEPKMQSKTKIYRGPLDR 508
 QY 58 PASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
 DB 509 PARLCSVDNDIEVTPPEISTAQPLLHNSVGS 542
 RESULT 20
 ADB84404
 ID ADB84404 standard; protein; 162 AA.
 XX ADB84404;
 XX 04-DEC-2003 (first entry)
 DT MSRV-1 associated protein/peptide sequence #11.
 DE multiple sclerosis; rheumatoid arthritis; gag; pol;
 XX reverse transcriptase; ribonuclease H.
 KW Unidentified;
 XX US2003039664-A1.
 PN

XX 27-FEB-2003.
 PD 26-NOV-1997; 97US-00979847.
 PF 26-NOV-1996; 96US-00756429.
 XX (PERR/) PERRON H.
 PA (BESE/) BESEME F.
 PA (BEDI/) BEDIN F.
 PA (PARA/) PARANHOS-BACCALA G.
 PA (KOMU/) KOMURIAN-PRADEL F.
 PA (JOLI/) JOLIVET-REYNAUD C.
 PA (MAND/) MANDRAND B.
 PA (GARS/) GARSON J A.
 PA (TUKE/) TUKE P W.
 XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
 PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
 XX WPI; 2003-512253/48.
 XX New isolated or purified nucleic acid associated with multiple sclerosis
 PT and/or rheumatoid arthritis, useful for detecting a virus associated with
 PT multiple sclerosis or rheumatoid arthritis in a biological sample.
 XX Claim 31; Page 81; 193pp; English.
 XX The invention relates to an isolated or purified nucleic acid from a
 CC virus associated with multiple sclerosis and/or rheumatoid arthritis,
 CC multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
 CC pol, gag or reverse transcriptase genes (or their fragments) encoding the
 CC proteins or defined peptides (including immunodominant peptides,
 CC antigenic peptides or conserved motifs). Also included are a process for
 CC detecting a virus associated with multiple sclerosis or rheumatoid
 CC arthritis in a biological sample, a nucleic acid probe for the detection
 CC of a virus associated with multiple sclerosis or rheumatoid arthritis, a
 CC primer for the amplification by polymerisation of a nucleic acid of a
 CC viral material associated with multiple sclerosis or rheumatoid
 CC arthritis, a polypeptide exhibiting an inhibitory activity on the
 CC proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
 CC and an antibody directed against the MSRV-1 virus obtained by
 CC immunologically reacting a human or animal body or cells with an
 CC immunogenic agent consisting of the antigenic polypeptide defined above.
 CC The nucleic acids are useful for detecting a biological sample a virus
 CC associated with multiple sclerosis or rheumatoid arthritis, or for
 CC detecting in a biological sample, the presence of or exposure to a virus
 CC associated with multiple sclerosis or rheumatoid arthritis. The present
 CC sequence is a claimed MSRV-associated sequence whose identity cannot be
 CC accurately determined. Note: The SEQ ID numbers for the sequences as
 CC displayed in the main body of the patent do not match the SEQ ID numbers
 CC in the sequence listing. Consequently those sequences mentioned in the
 CC claims may not be the sequences the authors intended to claim.
 XX Sequence 162 AA;
 SQ
 Query Match 75.3%; Score 344; DB 7; Length 162;
 Best Local Similarity 77.7%; Pred. No. 8.8e-33;
 Matches 73; Conservative 3; Mismatches 14; Indels 4; Gaps 1;
 QY 2 LPFLGPLAAIILLFPGPCIFNLLVNFVSSRIEAVK----LQMEPKMQSKTKIYRRPLDR 57
 DB 69 LPFLGPLAAIIFLLFPGPCIFNLFVSVSRIRIYVAVKQLQIVLQMEPKMQSKTKIYRGPLDR 128
 QY 58 PASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 91
 DB 129 PARLCSVDNDIEVTPPEISTAQPLLHNSVGS 162
 RESULT 21
 ABG15631
 ID ABG15631 standard; protein; 87 AA.
 XX

AC ABG15631;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15622.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX N-PSDB; AAS79818.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 45990; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 87 AA;
Query Match 66.0%; Score 301.5; DB 4; Length 87;
Best Local Similarity 71.0%; Pred. No. 4.8e-28;
Matches 66; Conservative 5; Mismatches 11; Indels 11; Gaps 2;
QY 3 PFLGFLAAIILLFPGCIFNLLVNFVSSRIEAVK- - - - - LQMEPFQMSKTKIYRRLDRP 58
Db 2 PALQPIH- - - - - FNPFIENLLVKFVSSRIKAVKLQIVLQMEPFQMSKTKIYHGLDRP 54
QY 59 ASPRSDVNDIKGTTPPEEISAAQPLLRNSAGSS 91
Db 55 ASPCSDVDDIEGTPPEEISTAQPLLRNSAGSS 87
RESULT 22
ABG27347

ID ABG27347 standard; protein; 87 AA.
XX
AC ABG27347;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27338.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX N-PSDB; AAS91534.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 57706; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 87 AA;
Query Match 66.0%; Score 301.5; DB 4; Length 87;
Best Local Similarity 71.0%; Pred. No. 4.8e-28;
Matches 66; Conservative 5; Mismatches 11; Indels 11; Gaps 2;
QY 3 PFLGFLAAIILLFPGCIFNLLVNFVSSRIEAVK- - - - - LQMEPFQMSKTKIYRRLDRP 58
Db 2 PALQPIH- - - - - FNPFIENLLVKFVSSRIKAVKLQIVLQMEPFQMSKTKIYHGLDRP 54
QY 59 ASPRSDVNDIKGTTPPEEISAAQPLLRNSAGSS 91
Db 55 ASPCSDVDDIEGTPPEEISTAQPLLRNSAGSS 87

RESULT 23
ABG08534
ID ABG08534 standard; protein; 144 AA.
XX AC
XX AC ABG08534;
XX DT 13-FEB-2002 (first entry)
XX XX
XX DE Novel human diagnostic protein #8525.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WC200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSB-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS72721.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 38893; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridization probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX XX
XX SQ Sequence 144 AA;
Query Match 65.2%; Score 298; DB 4; Length 144;
Best Local Similarity 90.9%; Pred. No. 2.3e-27;
Matches 60; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 26 VNFVSRIEAVKLOEPMQSKTKLYRRLDPASPRSDVNDIKGTPPEISAAQPLLRP 85
DB 79 VTDIGRIEAVKLOEPMQSKTKLYRRLDPASPRSDVNDIKGTPPEISAAQPLLRP 138
QY 86 NSAGSS 91
DB 139 NSAGSS 144

RESULT 24
ABR41810
ID ABR41810 standard; protein; 103 AA.
XX AC
XX AC ABR41810;
XX DT 02-JUN-2003 (first entry)
XX XX
XX DE Human DITHP growth/development-associated protein.
XX KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
XX KW cancer; cell proliferative disorder; autoimmune disorder;
XX KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
XX KW neurological disorder; gastrointestinal disorder; transport disorder;
XX KW connective tissue disorder; drug screening; proteome analysis;
XX KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
XX KW disease model; toxicological testing; transcript imaging; growth;
XX KW development.
XX OS Homo sapiens.
XX PN WO200297031-A2.
XX PD 05-DEC-2002.
XX PF 27-MAR-2002; 2002WO-US010056.
XX PR 28-MAR-2001; 2001US-0279619P.
XX PR 29-MAR-2001; 2001US-0280067P.
XX PR 29-MAR-2001; 2001US-0280068P.
XX PR 16-MAY-2001; 2001US-0291280P.
XX PR 17-MAY-2001; 2001US-0291829P.
XX PR 17-MAY-2001; 2001US-0291849P.
XX PR 19-JUN-2001; 2001US-0299428P.
XX PR 20-JUN-2001; 2001US-0299776P.
XX PR 20-JUN-2001; 2001US-0300001P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX PI Dufour GE, Hillman JB, Yu JY, Tuason O, Yap PE, Amshay SR;
XX PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
XX PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX PF: 2003-129518/12.
XX DR N-PSDB; ACC46747.
XX CC Novel human diagnostic and therapeutic polypeptide useful for identifying
CC test compound which specifically binds to a polypeptide encoded by human
CC diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX CC
XX PS Claim 27; SEQ ID NO 1345; 591pp; English.
XX CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to

CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which is associated with
CC growth and development. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 103 AA;

Query Match 64.1%; Score 293; DB 6; Length 103;
Best Local Similarity 77.2%; Pred. No. 6.1e-27;
Matches 61; Conservative 4; Mismatches 10; Indels 4; Gaps 1;
QY 17 FGPCIFNLLVNFVSSRIEAVK---LQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTP 72
DB 9 FNLICIFNLLVNFVSSRIETVKLQWLQMEPKMQSKTKIHRPLDQPASPCSDVNDIEGTP 68
QY 73 PEISAAQPLLRNSAGSS 91
DB 69 PEISTVQPLLCFNSAGSS 87

RESULT 25

AAW71067
ID AAW71067 standard; protein; 77 AA.

AC AAW71067;

DT 29-DEC-1998 (first entry)

DE Multiple sclerosis associated retrovirus protein 2.

XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;
KW env gene; rheumatoid arthritis-associated virus.

XX Multiple sclerosis associated retrovirus.

XX WO9823755-A1.

XX 04-JUN-1998.

XX 26-NOV-1997; 97WO-IB001482.

XX 26-NOV-1996; 96US-00756429.

XX (INMR) BIO MERIEUX.

XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;

XX WPI; 1998-322732/28.

XX N-PSDB; AAV43215.

XX New nucleic acid from retroviruses - useful for diagnosis, prevention and
PT treatment of, e.g. multiple sclerosis.

XX Disclosure; Page 183-184; 286pp; English.

XX The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) protein used in the method of the invention. The
CC invention provides complete or partial genomic sequences of the MSRV-1
CC pol gene, gag gene and env gene, and polypeptides encoded by these genes.
CC The invention also provides antibodies raised against the polypeptides.
CC The genomic sequences, polypeptides and antibodies are also claimed
CC useful for diagnosing infection by MS and rheumatoid arthritis-associated
CC viruses, and also for prevention and treatment of infection with these
CC viruses

XX
SQ Sequence 77 AA;

Query Match 62.8%; Score 287; DB 2; Length 77;
Best Local Similarity 77.9%; Pred. No. 2.2e-26;
Matches 60; Conservative 5; Mismatches 8; Indels 4; Gaps 1;
QY 19 PCIFNLLVNFVSSRIEAVK---LQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTP 74
DB 1 PCIFNLLVNFVSSRIETVKLQWLQMEPKMQSKTKIHRPLDQPASPCSDVNDIEGTP 60
QY 75 EISAAQPLLRNSAGSS 91
DB 61 EISTAQPLLCFNSAGSS 77

Search completed: April 19, 2004, 10:53:10
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 19, 2004, 10:33:58 ; Search time 59 Seconds
(without alignments)
2576.450 Million cell updates/sec

Title: US-10-069-883-1
Perfect score: 2871
Sequence: 1 MALPYHIFLFTVLLPSTLT.....PPEISAAQPLLRNSAGSS 538

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 2871 | 100.0 | 538 | 4 | AAB75138 |
| 2 | 2871 | 100.0 | 538 | 4 | AAB7652 |
| 3 | 2871 | 100.0 | 540 | 3 | ABP40980 |
| 4 | 2871 | 100.0 | 685 | 3 | ABP40988 |
| 5 | 2871 | 100.0 | 846 | 3 | ABP41050 |
| 6 | 2868 | 99.9 | 538 | 3 | AAV08622 |
| 7 | 2868 | 99.9 | 538 | 3 | AAV67313 |
| 8 | 2868 | 99.9 | 538 | 3 | AAE14540 |
| 9 | 2868 | 99.9 | 538 | 7 | ADC38777 |
| 10 | 2791.5 | 97.2 | 544 | 7 | ADP09243 |
| 11 | 2791 | 97.2 | 538 | 2 | AAW99345 |
| 12 | 2558 | 89.1 | 542 | 5 | AAE25054 |
| 13 | 2503.5 | 87.2 | 541 | 3 | AAAB08195 |
| 14 | 2490 | 86.7 | 542 | 3 | AAAB19069 |
| 15 | 2314 | 80.6 | 493 | 2 | AAW71068 |
| 16 | 2314 | 80.6 | 493 | 2 | AAW99553 |
| 17 | 2289 | 79.7 | 493 | 3 | ADP08401 |
| 18 | 1234.5 | 43.0 | 263 | 4 | AAAB18008 |
| 19 | 1234.5 | 43.0 | 263 | 4 | AAAB37043 |
| 20 | 1234.5 | 43.0 | 263 | 4 | AAAM30519 |
| 21 | 1234.5 | 43.0 | 263 | 4 | AAAB31812 |
| 22 | 1234.5 | 43.0 | 263 | 4 | AAAB22358 |
| 23 | 1234.5 | 43.0 | 263 | 4 | AAAM70185 |
| 24 | 1234.5 | 43.0 | 263 | 4 | AAAM57770 |
| 25 | 1234.5 | 43.0 | 263 | 4 | ABG51885 |

| | | | | | |
|----|--------|------|------|---|-----------|
| 26 | 1234.5 | 43.0 | 263 | 4 | AAAM05648 |
| 27 | 1234.5 | 43.0 | 263 | 5 | ABG39819 |
| 28 | 1178.5 | 41.0 | 633 | 4 | ABG28306 |
| 29 | 1174.5 | 40.9 | 368 | 4 | ABG28311 |
| 30 | 1171 | 40.8 | 531 | 5 | ABP63036 |
| 31 | 1126.5 | 39.2 | 446 | 4 | ABG07532 |
| 32 | 1126.5 | 39.2 | 446 | 4 | ABG20007 |
| 33 | 686 | 23.9 | 120 | 4 | AAAB75137 |
| 34 | 672 | 23.4 | 162 | 2 | AAW71069 |
| 35 | 672 | 23.4 | 162 | 2 | AAW99554 |
| 36 | 655 | 22.8 | 162 | 7 | ADB84404 |
| 37 | 585.5 | 20.4 | 567 | 2 | AAV71700 |
| 38 | 563.5 | 19.6 | 584 | 4 | ABG22871 |
| 39 | 563.5 | 19.6 | 584 | 4 | ABG19851 |
| 40 | 563.5 | 19.6 | 584 | 4 | ABG05606 |
| 41 | 554 | 19.3 | 1167 | 4 | ABG19838 |
| 42 | 554 | 19.3 | 1167 | 4 | ABG05605 |
| 43 | 554 | 19.3 | 1186 | 4 | ABG06819 |
| 44 | 554 | 19.3 | 1340 | 4 | ABG06717 |
| 45 | 534.5 | 18.6 | 587 | 4 | ABG12484 |
| 46 | 534.5 | 18.6 | 1121 | 4 | ABG18584 |
| 47 | 530.5 | 18.5 | 469 | 5 | ABG03726 |
| 48 | 506 | 17.6 | 538 | 5 | ABP69150 |
| 49 | 488 | 17.0 | 231 | 4 | ABG12209 |
| 50 | 484 | 16.9 | 433 | 2 | AAW36029 |
| 51 | 484 | 16.9 | 433 | 2 | AAW71060 |
| 52 | 484 | 16.9 | 433 | 7 | ADB84376 |
| 53 | 473 | 16.5 | 618 | 4 | ABE11379 |
| 54 | 473 | 16.5 | 618 | 4 | AAW40524 |
| 55 | 464 | 16.2 | 253 | 4 | AAW94312 |
| 56 | 423.5 | 14.8 | 638 | 4 | AAE70634 |
| 57 | 420 | 14.6 | 673 | 2 | AAW81573 |
| 58 | 419 | 14.6 | 640 | 2 | AAV26910 |
| 59 | 419 | 14.6 | 653 | 4 | AAV19887 |
| 60 | 419 | 14.6 | 654 | 4 | AAV19886 |
| 61 | 419 | 14.6 | 882 | 2 | AAW44501 |
| 62 | 419 | 14.6 | 882 | 2 | AAW44899 |
| 63 | 419 | 14.6 | 892 | 2 | AAW44903 |
| 64 | 418.5 | 14.6 | 632 | 2 | AAV26912 |
| 65 | 418.5 | 14.6 | 665 | 3 | AAV10045 |
| 66 | 418.5 | 14.6 | 665 | 3 | AAV54267 |
| 67 | 418.5 | 14.6 | 675 | 5 | ABE09157 |
| 68 | 418.5 | 14.6 | 742 | 2 | AAW44893 |
| 69 | 418.5 | 14.6 | 757 | 2 | AAW44894 |
| 70 | 418.5 | 14.6 | 773 | 2 | AAW44887 |
| 71 | 418.5 | 14.6 | 783 | 2 | AAW44895 |
| 72 | 418.5 | 14.6 | 788 | 2 | AAW44888 |
| 73 | 418.5 | 14.6 | 829 | 2 | AAW44889 |
| 74 | 418.5 | 14.6 | 880 | 2 | AAW44884 |
| 75 | 418.5 | 14.6 | 926 | 2 | AAW44883 |
| 76 | 417.5 | 14.5 | 662 | 2 | AAW68401 |
| 77 | 416 | 14.5 | 346 | 4 | AAE03533 |
| 78 | 416 | 14.5 | 438 | 4 | AAE03479 |
| 79 | 416 | 14.5 | 438 | 4 | AAE03436 |
| 80 | 416 | 14.5 | 438 | 5 | ABG63391 |
| 81 | 416 | 14.5 | 438 | 5 | ABG63390 |
| 82 | 415 | 14.5 | 640 | 2 | AAV26909 |
| 83 | 415 | 14.5 | 653 | 4 | AAV19883 |
| 84 | 415 | 14.5 | 654 | 4 | AAV19882 |
| 85 | 415 | 14.5 | 687 | 5 | ABG61828 |
| 86 | 414.5 | 14.4 | 665 | 2 | AAV17948 |
| 87 | 413.5 | 14.4 | 638 | 2 | AAW32098 |
| 88 | 413.5 | 14.4 | 638 | 2 | AAW70532 |
| 89 | 413.5 | 14.4 | 638 | 4 | AAV73287 |
| 90 | 413 | 14.4 | 653 | 6 | ABE82635 |
| 91 | 413 | 14.4 | 676 | 5 | ABE09158 |
| 92 | 413 | 14.4 | 1361 | 4 | ABG17354 |
| 93 | 413 | 14.4 | 1361 | 4 | ABG13631 |
| 94 | 412 | 14.4 | 192 | 6 | ABO10263 |
| 95 | 412 | 14.4 | 193 | 4 | AAU14098 |
| 96 | 411.5 | 14.3 | 538 | 4 | AAV35114 |
| 97 | 411.5 | 14.3 | 668 | 4 | ABE09159 |
| 98 | 411.5 | 14.3 | 672 | 5 | ABE09154 |

99 411.5 14.3 675 5 ABB09156 Abb09156 Murine le
100 411.5 14.3 675 5 ABB09155 Abb09155 Murine le

ALIGNMENTS

RESULT 1
AAB75138
ID AAB75138 standard; protein; 538 AA.
XX AC AAB75138;

XX 08-AUG-2001 (first entry)

XX HERV-W envelope protein G.

XX Human endogenous retrovirus; HERV-W; chromosome 7; env protein;
KW envelope protein; multiple sclerosis-related superantigen; vaccine;
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW antisense-therapy; autoimmune disorder.

XX Human endogenous retrovirus.

XX Key Location/Qualifiers
FH Region 271..317
FT /label= SU
FT /note= "surface protein"
FT Cleavage-site 317..318
FT /note= "SU-TM putative cleavage site"
FT Region 318..538
FT /label= TM
FT /note= "transmembrane subunit"

XX WO200131021-A1.

XX 03-MAY-2001.

XX 30-OCT-2000; 2000WO-EP010659.

XX 28-OCT-1999; 99EP-00402690.

XX (UYGE-) UNIV GENEVE.

XX Conrad B, Mach B;

XX WPI; 2001-316336/33.

XX N-PSDB; AAB20070.

XX New human retrovirus HERV-W ENV proteins/peptides having superantigen
PT activity useful for diagnosing and treating multiple sclerosis.

XX Claim 1; Fig 7; 94pp; English.

XX On the basis of the PBS t-RNA motif used for the classification of human
CC endogenous retrovirus (HERVs) the full length endogenous provirus which
CC has been located on the long arm of human chromosome 7 (7q21-22) has been
CC designated HERV-W. The present invention describes proteins or peptides
CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC have neuroprotective activity, and can be used in: vaccines; antisense-
CC therapy; and HERV-W SAG activity-inhibitors. (II) and encoding DNA/RNA are
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC disorders. (I) are also useful for identifying substances (and optionally
CC recovering) capable of binding to a retroviral superantigen associated
CC with MS, substances capable of blocking SAG activity and substances
CC capable of blocking transcription or translation of HERV-W retroviral
CC superantigen. A protein or peptide derived from (I), modified to be
CC devoid of SAG activity and being capable of generating an immune response
CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid
CC molecules encoding (I) are useful as vaccines against MS. Substances
CC capable of blocking SAG activity, capable of binding to a retroviral
CC superantigen associated with MS, or capable of blocking transcription or

CC translation of HERV-W retroviral superantigen for use in treating or
CC preventing MS, obtained using (I) are useful for the treatment and
CC prevention of MS. (I) and nucleic acids encoding them are useful for
CC diagnosing autoimmune disease. The present sequence represents the
CC specifically claimed envelope protein of HERV-W designated G

XX Sequence 538 AA;

Query Match 100.0%; Score 2871; DB 4; Length 538;
Best Local Similarity 100.0%; Pred. No. 2.1e-257;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSSPYQCEFLWMQRCGNIDAPYSLSKGTG 60
DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSSPYQCEFLWMQRCGNIDAPYSLSKGTG 60
QY 61 TPTAHTMPRCNCHSATLCMEHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTGMSDGGV 120
DB 61 TPTAHTMPRCNCHSATLCMEHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTGMSDGGV 120
QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLKLFHETLRTHTRLVSLFNTTLTGLHEVSA 180
DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLKLFHETLRTHTRLVSLFNTTLTGLHEVSA 180
QY 181 QNPTNCWICLPLNRPYVSIIPVPEQMNFTSTNTSVLVGPLYSNLEITHTSNLTCKVF 240
DB 181 QNPTNCWICLPLNRPYVSIIPVPEQMNFTSTNTSVLVGPLYSNLEITHTSNLTCKVF 240
QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVPPMTIY 300
QY 301 TEQDLYSVISKPRNKEVPILPFVIGAGVLGALGTGIGITTTQFYFKLSQELNGDMER 360
DB 301 TEQDLYSVISKPRNKEVPILPFVIGAGVLGALGTGIGITTTQFYFKLSQELNGDMER 360
QY 361 VADSLVTLQDQLNSLAAVLQNRALDILLTAERGTCFLGEECCYVNGSGIVTEKYKE 420
DB 361 VADSLVTLQDQLNSLAAVLQNRALDILLTAERGTCFLGEECCYVNGSGIVTEKYKE 420
QY 421 IRDRIQRRAEELRNTGPWLLSOMWPILPFLGFLAAIILLLLFGPCIFNLLVNFVSSRI 480
DB 421 IRDRIQRRAEELRNTGPWLLSOMWPILPFLGFLAAIILLLLFGPCIFNLLVNFVSSRI 480
QY 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPSAGSS 538
DB 481 EAVKLQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPSAGSS 538

RESULT 2

AAB67652
ID AAB67652 standard; protein; 538 AA.

XX AC AAB67652;

XX 29-MAY-2001 (first entry)

XX Amino acid sequence of a human endogenous retrovirus envelope protein.
KW Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion.

XX Human endogenous retrovirus.

XX WO200116171-A1.

XX 08-MAR-2001.

XX 01-SEP-2000; 2000WO-FR002429.

XX 01-SEP-1999; 99FR-00011141.

XX 15-SEP-1999; 99FR-00011793.

PA (INRM) BIO MERIEUX.
 FA (INRM) INST NAT SANTE & RECH MEDICALE.

PI Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
 XX
 XX
 XX WPI: 2004-226676/23.
 DR N-PSDB; AAF55630.
 XX

PT Detecting expression of human endogenous retrovirus envelope protein in
 cells of a tissue or culture, from its ability to induce syncytia.

XX
 XX Claim 1; Page 41-44; 57pp; French.

XX The present sequence represents a human endogenous retrovirus envelope
 CC protein. The specification describes a method for detecting expression of
 CC an envelope protein from a human endogenous retrovirus (HERV), in cells,
 CC of a tissue or culture. The method comprises detecting syncytia formation
 CC due to the fusogenic properties of the envelope protein. Envelope
 CC polypeptides and polynucleotides are used to produce therapeutic or
 CC prophylactic compositions, particularly for treatment of cancer, to
 CC correct defects in placental development (or other natural formation of
 CC other types of syncytia), and to promote adhesion of cells in grafts or
 CC cellular repair processes. Expression of sequences antisense to the
 CC polynucleotide are used to prevent formation of syncytia

XX Sequence 538 AA;

Query Match 100.0%; Score 2871; DB 4; Length 538;
 Best Local Similarity 100.0%; Pred. No. 2.1e-257;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSTSSPYQEFLLRMQPCGNIDAPYSRLSKGTP 60
 DB 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSTSSPYQEFLLRMQPCGNIDAPYSRLSKGTP 60
 QY 61 TPTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTCWYFTQTGMSDGGGV 120
 DB 61 TPTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTCWYFTQTGMSDGGGV 120
 QY 121 QDOAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLHETLRTHTLVSLENTLTGLHEVSA 180
 DB 121 QDOAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLHETLRTHTLVSLENTLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLGPVLSNLEITHTSNLTVCVKF 240
 DB 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLGPVLSNLEITHTSNLTVCVKF 240
 QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVLPMTIY 300
 DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVLPMTIY 300
 QY 301 TEQDLYSYVISKPRNKRPVILPFVIGAGVLGALGTGIGTITSTQFYKLSQELNGDME 360
 DB 301 TEQDLYSYVISKPRNKRPVILPFVIGAGVLGALGTGIGTITSTQFYKLSQELNGDME 360
 QY 361 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCCLFGERCCYVNGSGIVTEKVE 420
 DB 361 VADSLVTLQDQNSLAAVLQNRALDILLTAERGTCCLFGERCCYVNGSGIVTEKVE 420
 QY 421 IRDRIQRAEELRTGPGWLLSQMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 DB 421 IRDRIQRAEELRTGPGWLLSQMPWILPFLGPLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKLOMEPKMOSKTIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSGSS 538
 DB 481 EAVKLOMEPKMOSKTIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSGSS 538

RESULT 3

ABP40980

ID ABP40980 standard; protein; 540 AA.

XX

AC ABP40980;

XX 01-AUG-2002 (first entry)
 DT
 XX
 XX Human retroviral HERV-7q env peptide #3.
 DE
 XX
 XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 KW multiple sclerosis.
 XX
 XX Human endogenous retrovirus.
 OS
 XX
 XX WO9967395-A1.
 PN
 XX
 XX 29-DEC-1999.
 PD
 XX
 XX 23-JUN-1999; 99WO-PR001513.
 PF
 XX
 XX 23-JUN-1999; 98FR-00007920.
 PR
 XX
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX
 XX Alliel PM, Perin J, Rieger F;
 PI
 XX
 XX WPI: 2000-160587/14.
 DR

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 PT for diagnosis, treatment and prevention of autoimmune and neurological
 PT diseases.
 XX
 XX Claim 22; Fig 4; 225pp; French.
 PS
 XX
 XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention

XX Sequence 540 AA;

Query Match 100.0%; Score 2871; DB 3; Length 540;
 Best Local Similarity 100.0%; Pred. No. 2.1e-257;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSTSSPYQEFLLRMQPCGNIDAPYSRLSKGTP 60
 DB 3 MALPYHIFLFTVLLPSFTLTAPPCCRCWTSSTSSPYQEFLLRMQPCGNIDAPYSRLSKGTP 62
 QY 61 TPTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTCWYFTQTGMSDGGGV 120
 DB 63 TPTAHTMPRNCYHSATLCHMANTHYWTGKMINPSCPGGLGVTCWYFTQTGMSDGGGV 122
 QY 121 QDOAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLHETLRTHTLVSLENTLTGLHEVSA 180
 DB 123 QDOAREKHVKEVISQLTRVHGTSSTPYKGLDLSKLHETLRTHTLVSLENTLTGLHEVSA 182
 QY 181 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLGPVLSNLEITHTSNLTVCVKF 240
 DB 183 QNPTNCWICLPLNFRPVYSIPVPEQNNFSTEINTSVLGPVLSNLEITHTSNLTVCVKF 242
 QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVLPMTIY 300
 DB 243 SNTTYTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSESCFSLFVLPMTIY 302
 QY 301 TEQDLYSYVISKPRNKRPVILPFVIGAGVLGALGTGIGTITSTQFYKLSQELNGDME 360
 DB 303 TEQDLYSYVISKPRNKRPVILPFVIGAGVLGALGTGIGTITSTQFYKLSQELNGDME 362

QY 361 VADSLVTLODQNSLAADVQLQNRALDLTAERGCTCLFLGEECCYYVNSGIVTEKYKE 420
 Db 363 VADSLVTLODQNSLAADVQLQNRALDLTAERGCTCLFLGEECCYYVNSGIVTEKYKE 422
 QY 421 IRDRIQRAAEELRNTGPGWLLSQMPWILPFLGLAAIILLFPGCIFNLLVNFVSSRI 480
 Db 423 IRDRIQRAAEELRNTGPGWLLSQMPWILPFLGLAAIILLFPGCIFNLLVNFVSSRI 482
 QY 481 EAVKLQMEPRMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSAGSS 538
 Db 483 EAVKLQMEPRMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSAGSS 540

RESULT 4

ABP40988
 ID ABP40988 standard; protein; 685 AA.

XX AC ABP40988;
 XX DT 01-AUG-2002 (first entry)
 XX DE Human retroviral env ORF protein.
 XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 XX KM multiple sclerosis.
 XX OS Human endogenous retrovirus.
 XX PN WO9967395-A1.
 XX PD 29-DEC-1999.
 XX PF 23-JUN-1999; 99WO-FR001513.
 XX PR 23-JUN-1998; 98FR-00007920.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Alliel PM, Perin J, Rieger F;
 XX DR WPI; 2000-160587/14.

XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 XX PT for diagnosis, treatment and prevention of autoimmune and neurological
 XX PT diseases.
 XX PS Claim 22; Page 152-154; 225pp; French.
 CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer.
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention

Query Match 100.0%; Score 2871; DB 3; Length 685;
 Best Local Similarity 100.0%; Pred. No. 3e-257;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWRMORPGNIDAPSVRSLSKQTP 60
 Db 131 MALPYHIFLFTVLLPSFTLTAPPCCMTSSSPYQEFLLWRMORPGNIDAPSVRSLSKQTP 190

QY 61 TETAHTMPENYCHSATLCHNANTHYWTGWINPSCPGGLGVTVCTVYFTQTGMSDGGV 120
 Db 191 TETAHTMPENYCHSATLCHNANTHYWTGWINPSCPGGLGVTVCTVYFTQTGMSDGGV 250
 QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLBHETLRLTHRLVSLFNTTLTGLHEVSA 180
 Db 251 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLBHETLRLTHRLVSLFNTTLTGLHEVSA 310
 QY 181 QNPTNCWICLPLNFRPVVSIIPVEQNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKF 240
 Db 311 QNPTNCWICLPLNFRPVVSIIPVEQNNFSTEINTTSLVGLPLVSNLEITHTSNLTVCVKF 370
 QY 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCINGSSBSCMCFSLFVPPMTIY 300
 Db 371 SNTTYTNSQCIRWVTPPTQIVCLPSGIFPVCGTSAYRCINGSSBSCMCFSLFVPPMTIY 430
 QY 301 TEQDLYSVYLSKPRNKRVPILPEVIGAGVLGALGTGIGITTSQFYFKLSQELNGDMER 360
 Db 431 TEQDLYSVYLSKPRNKRVPILPEVIGAGVLGALGTGIGITTSQFYFKLSQELNGDMER 490
 QY 361 VADSLVTLODQNSLAADVQLQNRALDLTAERGCTCLFLGEECCYYVNSGIVTEKYKE 420
 Db 491 VADSLVTLODQNSLAADVQLQNRALDLTAERGCTCLFLGEECCYYVNSGIVTEKYKE 550
 QY 421 IRDRIQRAAEELRNTGPGWLLSQMPWILPFLGLAAIILLFPGCIFNLLVNFVSSRI 480
 Db 551 IRDRIQRAAEELRNTGPGWLLSQMPWILPFLGLAAIILLFPGCIFNLLVNFVSSRI 610
 QY 481 EAVKLQMEPRMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSAGSS 538
 Db 611 EAVKLQMEPRMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNSSAGSS 668

RESULT 5

ABP41050
 ID ABP41050 standard; protein; 846 AA.

XX AC ABP41050;
 XX DT 01-AUG-2002 (first entry)
 XX DE HERV-7q env protein #1.
 XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 XX KW multiple sclerosis.
 XX OS Human endogenous retrovirus.
 XX PN WO9967395-A1.
 XX PD 29-DEC-1999.
 XX PF 23-JUN-1999; 99WO-FR001513.
 XX PR 23-JUN-1998; 98FR-00007920.
 XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Alliel PM, Perin J, Rieger F;
 XX DR WPI; 2000-160587/14.

XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q, used
 XX PT for diagnosis, treatment and prevention of autoimmune and neurological
 XX PT diseases.

XX PS Example 4; Fig 19; 225pp; French.

XX CC The present invention relates to new nucleic acid sequences of human
 XX CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 XX CC Regulatory elements associated with HERV-7q may alter expression of other
 XX CC genes (even remote genes) on the same chromosome, inducing immunological
 XX CC and/or neurological changes (which may be pathological or protective/

CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention
 CC
 XX Sequence 846 AA;

Query Match 100.0%; Score 2871; DB 3; Length 846;
 Best Local Similarity 100.0%; Pred. No. 4.1e-257;
 Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLWMQRPNGNDAPSYSLSKGP | 60 |
| DB | 291 | MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLWMQRPNGNDAPSYSLSKGP | 350 |
| QY | 61 | TFTATHMPRNCYHSATLCMEANTHYWTGKMINPSCPGGLGVTVVCTYFTQTGMSDGGV | 120 |
| DB | 351 | TFTATHMPRNCYHSATLCMEANTHYWTGKMINPSCPGGLGVTVVCTYFTQTGMSDGGV | 410 |
| QY | 121 | QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETLRLVSLFNTLTGLHEVSA | 180 |
| DB | 411 | QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETLRLVSLFNTLTGLHEVSA | 470 |
| QY | 181 | QNPTNCWICLPLNFRPYVSIIPVPEQWNNFSTEINTSVLGPLVSNLEIHTNSLTCVKF | 240 |
| DB | 471 | QNPTNCWICLPLNFRPYVSIIPVPEQWNNFSTEINTSVLGPLVSNLEIHTNSLTCVKF | 530 |
| QY | 241 | SNNTYTTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCNGSSSCFSLFVPPMTIY | 300 |
| DB | 531 | SNNTYTTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCNGSSSCFSLFVPPMTIY | 590 |
| QY | 301 | TEQDLYSVISKPRNKRVPILPFVIGAGVLCALGTGGITTSQFYKLSQELNGDMER | 360 |
| DB | 591 | TEQDLYSVISKPRNKRVPILPFVIGAGVLCALGTGGITTSQFYKLSQELNGDMER | 650 |
| QY | 361 | VADSLVTLDQOLNSLAAVLQNRALDILLTAERGCTCLFGECCYVYVNSGIVTEKVK | 420 |
| DB | 651 | VADSLVTLDQOLNSLAAVLQNRALDILLTAERGCTCLFGECCYVYVNSGIVTEKVK | 710 |
| QY | 421 | IRDRIQRAAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI | 480 |
| DB | 711 | IRDRIQRAAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI | 770 |
| QY | 481 | EAVKLOMEPKQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS | 538 |
| DB | 771 | EAVKLOMEPKQSKTKIYRPLDRPASPRSDVNDIKGTPPEEISAAQPLLRPNAGSS | 828 |

RESULT 6
 AAY08622
 ID AAY08622 standard; protein; 538 AA.
 XX
 AC AAY08622;
 XX
 DT 10-AUG-1999 (first entry)
 XX
 DE Human secreted protein AJ172_2.

Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
 bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
 cell proliferation; cell differentiation; suppressor; tumour inhibition;
 haematopoiesis regulator; activin; inhibin; chemotactic; chemokine;
 haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
 cadherin; tumour invasion suppressor; gene therapy; tissue growth.

OS Homo sapiens.
 XX
 PN WO9926972-A1.
 XX

PD 03-JUN-1999.
 XX
 PF 17-NOV-1998; 98WO-US024614.
 XX
 PR 21-NOV-1997; 97US-00976110.
 PR 18-MAY-1998; 98US-00080478.
 PR 20-OCT-1998; 98US-00175928.
 XX
 PA (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M;
 XX
 DR WPI; 1999-357813/30.
 DR N-PSDB; AAX77526.
 XX
 XX New polynucleotides encoding secreted proteins.
 XX
 PS Claim 14a; Page 101-103; 142pp; English.
 XX
 CC This invention describes novel human secreted proteins encoded by
 CC polynucleotides isolated from human adult testes, adult brain, adult
 CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
 CC libraries. The products of the invention are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional
 CC activity, cytokine and cell proliferation/differentiation activity,
 CC immune stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, and tumour
 CC inhibition activity. The polynucleotides are also stated to be useful for
 CC gene therapy
 XX
 SQ Sequence 538 AA;

Query Match 99.9%; Score 2868; DB 2; Length 538;
 Best Local Similarity 99.8%; Pred. No. 4e-257; 0; Indels 0; Gaps 0;
 Matches 537; Conservative 1; Mismatches 0;

| | | | |
|----|-----|--|-----|
| QY | 1 | MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWMQRPNGNDAPSYSLSKGP | 60 |
| DB | 1 | MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSPYQEFLLWMQRPNGNDAPSYSLSKGP | 60 |
| QY | 61 | TFTATHMPRNCYHSATLCMEANTHYWTGKMINPSCPGGLGVTVVCTYFTQTGMSDGGV | 120 |
| DB | 61 | TFTATHMPRNCYHSATLCMEANTHYWTGKMINPSCPGGLGVTVVCTYFTQTGMSDGGV | 120 |
| QY | 121 | QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETLRLVSLFNTLTGLHEVSA | 180 |
| DB | 121 | QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLEHETLRLVSLFNTLTGLHEVSA | 180 |
| QY | 181 | QNPTNCWICLPLNFRPYVSIIPVPEQWNNFSTEINTSVLGPLVSNLEIHTNSLTCVKF | 240 |
| DB | 181 | QNPTNCWICLPLNFRPYVSIIPVPEQWNNFSTEINTSVLGPLVSNLEIHTNSLTCVKF | 240 |
| QY | 241 | SNNTYTTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCNGSSSCFSLFVPPMTIY | 300 |
| DB | 241 | SNNTYTTNSQCIRWVTPPTQIVCLPSGIFFCVGTSAVRCNGSSSCFSLFVPPMTIY | 300 |
| QY | 301 | TEQDLYSVISKPRNKRVPILPFVIGAGVLCALGTGGITTSQFYKLSQELNGDMER | 360 |
| DB | 301 | TEQDLYSVISKPRNKRVPILPFVIGAGVLCALGTGGITTSQFYKLSQELNGDMER | 360 |
| QY | 361 | VADSLVTLDQOLNSLAAVLQNRALDILLTAERGCTCLFGECCYVYVNSGIVTEKVK | 420 |
| DB | 361 | VADSLVTLDQOLNSLAAVLQNRALDILLTAERGCTCLFGECCYVYVNSGIVTEKVK | 420 |
| QY | 421 | IRDRIQRAAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI | 480 |
| DB | 421 | IRDRIQRAAEELRNTGPGWLLSQWMPWILPFLGPLAAIILLFGPCIFNLLVNFVSSRI | 480 |

QY 481 EAVKQMEPKMOSKTKIYRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538
 DB 481 EAVKQMEPKMOSKTKIYRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538

RESULT 7
 ID AAY67313 standard; protein; 538 AA.
 AC AAY67313;
 XX
 XX
 DT 11-APR-2000 (first entry)
 DE Human secreted protein AJ172_2 amino acid sequence.
 KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
 KW placental pathology; metastasis inhibition; nutritional activity;
 KW immune stimulator; haematopoiesis regulator; tissue growth;
 KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115.
 XX
 OS Homo sapiens.
 XX
 XX WO9960020-A1.
 XX
 XX 25-NOV-1999.
 XX
 XX 17-MAY-1999; 99WO-US010915.
 XX
 XX 18-MAY-1998; 98US-00080478.
 XX
 XX 20-OCT-1998; 98US-00175928.
 XX
 XX (GEMY) GENETICS INST INC.
 XX
 XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 XX Merberg D, Mi S, Treacy M;
 XX
 XX WPI; 2000-116311/10.
 XX
 XX N-PSDB; AAZ59468.
 XX
 XX New polynucleotides encoding secreted cDNA libraries, used to develop
 XX products for the diagnosis and treatment of neoplastic disease.
 XX
 XX Claim 15; Page 108-110; 149pp; English.
 XX
 XX This is the human secreted protein AJ172_2 amino acid sequence, the
 XX polynucleotide encoding AJ172_2 was obtained from a human adult testes
 XX cDNA library. The invention relates to secreted human and murine
 XX proteins. The polynucleotides and proteins are predicted to have
 XX biological activities which would make them suitable for treating,
 XX preventing or ameliorating medical conditions in humans and animals.
 XX Detection of the levels of the proteins can be used for the diagnosis of
 XX e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate
 XX the expression or function of the proteins may be used for treating a
 XX neoplastic disease and inhibiting metastasis. Other suggested activities
 XX include nutritional activity (e.g. in feeds), cytokine and cell
 XX proliferation/differentiation activity, immune stimulating (e.g. as
 XX vaccines) or suppressing activity, haematopoiesis regulating activity,
 XX tissue growth activity, activin/inhibin activity,
 XX chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 XX receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 XX invasion suppressor activity, and tumour inhibition activity. The
 XX polynucleotide sequences are also stated to be useful for gene therapy
 XX
 XX Sequence 538 AA;
 XX
 XX Query Match 99.9%; Score 2868; DB 3; Length 538;
 XX Best Local Similarity 99.8%; Pred. No. 4e-257;
 XX Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX 1 MALPHIPLFTVLLPSFTLTAPPCKRMTSSSPYOEFLLRMORPCGNIDAPYSRLSKGTP 60
 XX
 XX 1 MALPHIPLFTVLLPSFTLTAPPCKRMTSSSPYOEFLLRMORPCGNIDAPYSRLSKGTP 60

QY 61 TFTAHTMPNCHXSATLCMHANTHYWTGKWINPSPGGLGVTVCWTYFTQTGMSGGGV 120
 DB 61 TFTAHTMPNCHXSATLCMHANTHYWTGKWINPSPGGLGVTVCWTYFTQTGMSGGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKIHETLRHTRLVSLFNTTLTGLHEVSA 180
 DB 121 QDQAREKHVKEVISQLTRVHGTSSPYKGLDLSKIHETLRHTRLVSLFNTTLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPYVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTHTNLTCVKF 240
 DB 181 QNPTNCWICLPLNFRPYVISIPVPEQNNFSTEINTSVLGPVLSNLEIHTHTNLTCVKF 240
 QY 241 SNTYTTNSOCIRWVTPPTQIVCLPSGIFPVCGTSAVRCCLNGSSSESCFSLFVPPMTIY 300
 DB 241 SNTYTTNSOCIRWVTPPTQIVCLPSGIFPVCGTSAVRCCLNGSSSESCFSLFVPPMTIY 300
 QY 301 TEQDLYSVISKPRNKRVPILPFVIGAGVLGALTGTGIGTITSTQFYKLSQELNGDMER 360
 DB 301 TEQDLYSVISKPRNKRVPILPFVIGAGVLGALTGTGIGTITSTQFYKLSQELNGDMER 360
 QY 361 VADSLVTLQDQLNSLAAVLQNRALDILLTAERGGTCLFLGEECCYYVNSGIVTEKVE 420
 DB 361 VADSLVTLQDQLNSLAAVLQNRALDILLTAERGGTCLFLGEECCYYVNSGIVTEKVE 420
 QY 421 IRDRIQRAEELRNTGFWGLLSQMPWILPFLGFLAAIILLLFGPCIFNLLVNFVSSRI 480
 DB 421 IRDRIQRAEELRNTGFWGLLSQMPWILPFLGFLAAIILLLFGPCIFNLLVNFVSSRI 480
 QY 481 EAVKQMEPKMOSKTKIYRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538
 DB 481 EAVKQMEPKMOSKTKIYRPLDRPASPRSDVNDIKGTPEEISAAQPLLRNSAGSS 538

RESULT 8
 ID AAE14540 standard; protein; 538 AA.
 XX
 XX AAE14540;
 AC
 XX
 XX 07-MAY-2002 (first entry)
 XX
 XX Human syncytin protein.
 DE
 XX
 XX Human; syncytin; pre-eclampsia; gestational trophoblast disorder;
 XX choriocarcinoma; hydatiform mole; placental site tumour; abortion;
 XX envelope gene; human endogenous defective retrovirus; HERV-W.
 OS
 XX Homo sapiens.
 XX
 XX WO200204578-A2.
 XX
 XX 17-JAN-2002.
 XX
 XX 09-JUL-2001; 2001WO-US021719.
 XX
 XX 07-JUL-2000; 2000US-0216657P.
 XX
 XX (GEMY) GENETICS INST INC.
 XX
 XX Keith JC, Mccoy JM, Mi S;
 XX WPI; 2002-171727/22.
 XX
 XX N-PSDB; AAD24195.
 XX
 XX Identifying a compound for treating a subject with or at risk of
 XX developing pre-eclampsia, comprises determining whether the expression or
 XX activity of syncytin in the cell is modulated in the presence of a test
 XX compound.
 XX
 XX Disclosure; Page 42-43; 43pp; English.
 XX
 XX The invention relates to identifying compounds which are modulators of

CC syncytin expression. The syncytin modulators are useful in diagnosis and
 CC treatment of preclampsia and gestational trophoblast disorders (e.g.
 CC choriocarcinoma, hydatiform mole, placental site tumour and missed/
 CC incomplete abortion). Syncytin is a human gene derived from the envelope
 CC gene of human endogenous defective retrovirus, HERV-W. The present
 CC invention is based partly on the discovery that syncytin expression is
 CC dramatically reduced in preclampsia, and is also mis-localised to the
 CC apical syncytiotrophoblast membrane. The present sequence is human
 CC syncytin protein
 XX
 XX Sequence 538 AA;
 SQ
 Query Match 99.9%; Score 2868; DB 5; Length 538;
 Best Local Similarity 99.8%; Pred. No. 4e-257;
 Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MALPHYHFLFTVLLPSFTLTAPPCCRCMTSSSPQEFLLWRMQRPCNDAPSYRLSKGTP 60
 DB 1 MALPHYHFLFTVLLPSFTLTAPPCCRCMTSSSPQEFLLWRMQRPCNDAPSYRLSKGTP 60
 QY 61 TFTAHTMPCRYHATLCMHANTHYTGKMINPCSCGGLGVTCWTVFTCTGMSDGGV 120
 DB 61 TFTAHTMPCRYHATLCMHANTHYTGKMINPCSCGGLGVTCWTVFTCTGMSDGGV 120
 QY 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRLTRVLSLFTNTLTGLHEVSA 180
 DB 121 QDQAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRLTRVLSLFTNTLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNRPYYSIVPPEONNPFSTEINTSVLGPLVSNLEIHTSNTLCVKF 240
 DB 181 QNPTNCWICLPLNRPYYSIVPPEONNPFSTEINTSVLGPLVSNLEIHTSNTLCVKF 240
 QY 241 SNTTNTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSCMCFSLFPPMTIY 300
 DB 241 SNTTNTNSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSCMCFSLFPPMTIY 300
 QY 301 TEODLYSVISKPRNRPVILPFVIGAGVLGALGTGIGGITTSTQPYKLSQELNGDWER 360
 DB 301 TEODLYSVISKPRNRPVILPFVIGAGVLGALGTGIGGITTSTQPYKLSQELNGDWER 360
 QY 361 VADSLVTLQDQLNSLAADVQLNRRALDLTAERGTCCLFLGEECCYVYVQSGIVTEKVE 420
 DB 361 VADSLVTLQDQLNSLAADVQLNRRALDLTAERGTCCLFLGEECCYVYVQSGIVTEKVE 420
 QY 421 TRDRTORABEURNTPGWLGLSQWMPWILPFLGLPLAAILLLLPGPCIFNLLNVFSRSRI 480
 DB 421 TRDRTORABEURNTPGWLGLSQWMPWILPFLGLPLAAILLLLPGPCIFNLLNVFSRSRI 480
 QY 481 EAVKLQMPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 538
 DB 481 EAVKLQMPKQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRPNAGSS 538
 RESULT 9
 ID ADC38777 standard; protein; 538 AA.
 XX
 AC ADC38777;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human secreted protein #63.
 KW immune disorder; severe combined immunodeficiency; SCID;
 KW autoimmune disorder; multiple sclerosis; systemic lupus erythematosus;
 KW rheumatoid arthritis; allergic reaction; asthma; myeloid cell deficiency;
 KW lymphoid cell deficiency; osteoporosis; osteoarthritis;
 KW peripheral nervous system disease; peripheral neuropathy;
 KW Alzheimer's disease; Parkinson's disease; coagulation disorder;
 KW inflammatory disease; systemic inflammatory response syndrome; SIRS;
 KW ischaemia-reperfusion injury; Crohn's disease; anaphylaxis;
 KW hypersensitivity; regeneration; neural cell proliferation; fertility;
 KW tumour; chemokine; human; secreted protein.

XX OS Homo sapiens.
 XX US2002193567-A1.
 XX 19-DEC-2002.
 XX 02-APR-2002; 2002US-00114893.
 XX 11-AUG-1995; 95US-00514014.
 PR 05-APR-1996; 96US-00628364.
 PR 19-APR-1996; 96US-00635311.
 PR 07-JUN-1996; 96US-00659224.
 PR 17-JUN-1996; 96US-00664596.
 PR 09-JUL-1996; 96US-00677231.
 PR 26-JUL-1996; 96US-00686878.
 PR 23-AUG-1996; 96US-00701819.
 PR 27-SEP-1996; 96US-00721488.
 PR 27-SEP-1996; 96US-00721798.
 PR 27-SEP-1996; 96US-00721923.
 PR 27-SEP-1996; 96US-00721926.
 PR 25-OCT-1996; 96US-00738367.
 PR 30-OCT-1996; 96US-00739775.
 PR 13-JAN-1997; 97US-00783395.
 PR 10-APR-1997; 97US-00833823.
 PR 02-JUN-1997; 97US-00867677.
 PR 05-SEP-1997; 97US-00924838.
 PR 06-OCT-1999; 99US-00413232.
 XX (GEMY) GENETICS INST INC.
 XX Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Bowman MR, Spaulding V, Carlin-Duckett M;
 PI Kelleher K;
 XX WPI: 2003-657236/62.
 DR N-PSDB; ADC38776.
 XX
 PT Proteins AZ3021 encoded by clone AZ3021 from human adult colon, and
 PT BD12716 encoded by clone BD12716 from human fetal kidney cDNA library,
 PT useful for treating e.g. multiple sclerosis and rheumatoid arthritis.
 XX
 PS Disclosure; SEQ ID NO 135; 412bp; English.
 XX
 CC The invention relates to a protein comprising fully defined AZ302 1
 CC protein or BD127 1 6 protein. The polynucleotides are useful for
 CC expressing recombinant proteins for analysis and are also useful as
 CC chromosome markers or tags to identify chromosomes or to map related gene
 CC positions. The proteins are useful as amino acid supplement, carbon
 CC source, nitrogen source and carbohydrate source. The proteins are useful
 CC for treating various immune deficiencies and disorders (e.g. severe
 CC combined immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus, rheumatoid arthritis), allergic
 CC reactions (e.g. asthma), myeloid or lymphoid cell deficiencies,
 CC osteoporosis or osteoarthritis, peripheral nervous system diseases (e.g.
 CC peripheral neuropathy, Alzheimer's disease, Parkinson's disease),
 CC coagulation disorders, inflammatory diseases (e.g. systemic inflammatory
 CC response syndrome (SIRS), ischaemia-reperfusion injury, Crohn's disease),
 CC anaphylaxis and hypersensitivity. Proteins are also useful for inducing
 CC tumour immunity, for inducing bone, cartilage, tendon, ligament and/or
 CC nerve growth or regeneration, for proliferating neural cells and for
 CC regenerating nerve and brain tissue, for inducing fertility and for
 CC inhibiting tumour growth. Proteins are also useful as chemokine for
 CC mammalian cells (e.g., monocytes, fibroblasts, neutrophils), and also
 CC useful as inhibitors of receptor/ligand interactions. The present
 CC sequence represents the amino acid sequence of a human secreted protein.
 XX
 SQ Sequence 538 AA;
 Query Match 99.9%; Score 2868; DB 7; Length 538;
 Best Local Similarity 99.8%; Pred. No. 4e-257;
 Matches 537; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPNGNIDAPSYRSLSGKTP 60
 Db 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPNGNIDAPSYRSLSGKTP 60
 QY 61 TFTAHTHMPNCVHSATLCHWANTHWTKMNPSCPGGLGVTVCTYFTQTMGSDGGV 120
 Db 61 TFTAHTHMPNCVHSATLCHWANTHWTKMNPSCPGGLGVTVCTYFTQTMGSDGGV 120
 QY 121 QDOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLTHTRLVSLFNLTGLHEVSA 180
 Db 121 QDOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLTHTRLVSLFNLTGLHEVSA 180
 QY 181 QNPTNCWICLPLNFRPYVSPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 240
 Db 181 QNPTNCWICLPLNFRPYVSPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 240
 QY 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVRCINGSSSMCFSLFVPPMTIY 300
 Db 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVRCINGSSSMCFSLFVPPMTIY 300
 QY 301 TEQDLYSYVIS-KPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSOELNGDME 360
 Db 301 TEQDLYSYVIS-KPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSOELNGDME 360
 QY 361 VADSLVTLQDQLNSLAADVQLNRRALDILLTAERGTCCLFLGEECCVYVNSGIVTEKYK 420
 Db 361 VADSLVTLQDQLNSLAADVQLNRRALDILLTAERGTCCLFLGEECCVYVNSGIVTEKYK 420
 QY 421 IRDRIQRABEELRNTGPGWLLSQMMPWILPFLGPLAAIILLFPGCIPNLLVNFVSSRI 480
 Db 421 IRDRIQRABEELRNTGPGWLLSQMMPWILPFLGPLAAIILLFPGCIPNLLVNFVSSRI 480
 QY 481 EAVKLOMEPKMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 538
 Db 481 EAVKLOMEPKMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

RESULT 10

AD09243
 ID ADE09243 standard; protein; 544 AA.

AC ADE09243;

XX 29-JAN-2004 (first entry)

DE Novel protein-related contig polypeptide sequence #309.

XX novel Gene; novel protein; tissue marker; molecular weight marker;

KW chromosome marker; genetic disorder; contig.

XX Unidentified.

OS WO2003054152-A2.

PN 03-JUL-2003.

XX 10-DEC-2002; 2002WO-US039555.

XX 10-DEC-2001; 2001US-0339739P.

PR 11-DEC-2001; 2001US-0339453P.

PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.

PR 12-APR-2002; 2002US-0372381P.

PR 12-APR-2002; 2002US-0372615P.

PR 22-APR-2002; 2002US-00128558.

PR 24-APR-2002; 2002US-0376045P.

XX (HYS-) HYSQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Wang G, Zhou P, Dmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX

DR WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization of therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.

PS Disclosure; SEQ ID NO 2787; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence was used in the
 CC exemplification of the invention.

XX Sequence 544 AA;

Query Match 97.2%; Score 2791.5; DB 7; Length 544;

Best Local Similarity 97.8%; Pred. No. 5.2e-250; Indels 1; Gaps 1;

Matches 527; Conservative 2; Mismatches 9;

QY 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPNGNIDAPSYRSLSGKTP 60

Db 6 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLRMQRPNGNIDAPSYRSLSGKTP 65

QY 61 TFTAHTHMPNCVHSATLCHWANTHWTKMNPSCPGGLGVTVCTYFTQTMGSDGGV 120

Db 66 TFTAHTHMPNCVHSATLCHWANTHWTKMNPSCPGGLGVTVCTYFTQTMGSDGGV 125

QY 121 QDOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLTHTRLVSLFNLTGLHEVSA 180

Db 126 QDOAREKHVKEVISQLTRGHSTSPYKGLVLSKLHETLTHTRLVSLFNLTGLHEVSA 185

QY 181 QNPTNCWICLPLNFRPYVSPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 240

Db 186 QNPTNCWICLPLNFRPYVSPVPEQNNFSTEINTSVLVGLVSNLEITHTSNLTCKVF 245

QY 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVRCINGSSSMCFSLFVPPMTIY 300

Db 246 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVRCINGSSSMCFSLFVPPMTIY 305

QY 301 TEQDLYSYVIS-KPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSOELNGDME 359

Db 306 TEQDLYSYVIS-KPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSOELNGDME 365

QY 360 RVADSLVTLQDQLNSLAADVQLNRRALDILLTAERGTCCLFLGEECCVYVNSGIVTEKYK 419

Db 366 QVADSLVTLQDQLNSLAADVQLNRRALDILLTAERGTCCLFLGEECCVYVNSGIVTEKYK 425

QY 420 EIRDRIQRABEELRNTGPGWLLSQMMPWILPFLGPLAAIILLFPGCIPNLLVNFVSSR 479

Db 426 EIRDRIQRABEELRNTGPGWLLSQMMPWILPFLGPLAAIILLFPGCIPNLLVNFVSSR 485

QY 480 IEAVKLOMEPKMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

Db 486 IEAVKLOMEPKMOSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 544

RESULT 11

AAW99345

ID AAW99345 standard; protein; 538 AA.

XX AAW99345;

AC AAW99345;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W ORF1.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

KW multiple sclerosis; rheumatoid polyarthritides; insulin-dependent diabetes;

KW disseminated lupus erythematosus; pregnancy; chromosomal marker.

XX Human endogenous retrovirus.

OS WO9902696-A1.

XX 21-JAN-1999.

PD 06-JUL-1998; 98WO-FR001442.

XX 07-JUL-1997; 97FR-00008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond J, Bouton O, Mandrand B, Mallet F;

XX WPI; 1999-120897/10.

DR New nucleic acid sequences from human endogenous retrovirus-W - expressed

PT exclusively in placenta and useful in diagnosis and therapy of autoimmune

PT disease, and abnormal or failed pregnancy.

XX Example 8; Fig 6; 106pp; French.

XX This sequence represents the open reading frame 1 encoding the env

CC protein from the human endogenous retrovirus (HERV) W (AA25665). Nucleic

CC acids, their fragments or peptides encoded by them derived from the HERV-

CC W genome are markers of autoimmune disease (e.g. multiple sclerosis,

CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-

CC dependent diabetes and related pathologies) and of abnormal or

CC unsuccessful pregnancy and can be used as chromosomal markers for

CC susceptibility to these conditions, or proximity markers of genes

CC associated with this susceptibility

XX Sequence 538 AA;

SQ

Query Match 97.2%; Score 2791; DB 2; Length 538;

Best Local Similarity 97.2%; Pred. No. 5.7e-250;

Matches 523; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 MALPHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLLMRMORGNIDAPSYRLSKGTP 60
 DB 1 MGLPWHIFLCSVLSFCFTLTAPPCCRCMTSSSPHPEFLMRMORGNIDAPSYRLSKGTP 60
 QY 61 TFTAHTMPCNYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTMGSDGGV 120
 DB 61 TFTAHTMPCNYHSATLCMHANTHYWTGKMINPSCPGGLGVTVCTWYFTQTMGSDGGV 120
 QY 121 QDQAREKHVKEVISQLTGVHGTSSPYKGLDLSKLHETLRTHLVLSLFTNTLTGLHEVSA 180
 DB 121 QDQAREKHVKEVISQLTGVHGTSSPYKGLDLSKLHETLRTHLVLSLFTNTLTGLHEVSA 180
 QY 181 QNPTNCWICLFLNRPYYSIPVPSQWNNFSTEINTSVLGPLVSNLEITHSLNLTCKVF 240
 DB 181 QNPTNCWICLFLNRPYYSIPVPSQWNNFSTEINTSVLGPLVSNLEITHSLNLTCKVF 240
 QY 241 SNTTYTNSQCIRWTPPTQIVCLPSGTFVCGTSAYRCLNGSSSCFSLFPPMTIY 300
 DB 241 SNTTYTNSQCIRWTPPTQIVCLPSGTFVCGTSAYRCLNGSSSCFSLFPPMTIY 300
 QY 301 TEQDLYSVISKPRNKRVPILPFVIGALGTGIGITTSQFYKLSQELNGDMER 360
 DB 301 TEQDLYSVISKPRNKRVPILPFVIGALGTGIGITTSQFYKLSQELNGDMER 360
 QY 361 VADSLVTLQDQINSAAVVLQNRALDLTLTAERGTCFLGEECCYVNSGIVTEKVE 420
 DB 361 VADSLVTLQDQINSAAVVLQNRALDLTLTAERGTCFLGEECCYVNSGIVTEKVE 420
 QY 421 IRDRIQRAEELRNTGPGWLLSQWNPWILPFLGLAAIILLLLFGPCITFLLNVFVSRI 480
 DB 421 IPDRIQRAEELRNTGPGWLLSRWNPWILPFLGLAAIILLLLFGPCITFLLNVFVSRI 480
 QY 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

Db 481 EAVKLQMEPKMQSKTKIYRRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAGSS 538

RESULT 12

AAE25054

ID AAE25054 standard; protein; 542 AA.

XX AAE25054;

AC AAE25054;

XX 30-OCT-2002 (first entry)

XX Human EMERY-2 protein.

XX Human; embryogenesis associated protein; AIDS; reproductive disorder;

XX infertility; endometriosis; endometrial tumour; inflammatory disorder;

XX autoimmune disorder; acquired immune deficiency syndrome; transgenic;

XX ovarian tumour; contact dermatitis; placenta disorder; preclampsia;

XX EMERY-2; allergy; gene therapy.

XX Homo sapiens.

XX Key

XX Peptide

XX Location/Qualifiers

XX 1..20

XX /label= Signal_peptide

XX Protein

XX 21..542

XX /note= "Mature EMERY-2 protein"

XX Domain

XX 263..484

XX /note= "ENV polypeptide (coat polypeptide) domain"

XX Domain

XX 455..473

XX /note= "Transmembrane domain"

XX WO200248362-A2.

XX 20-JUN-2002.

XX 14-NOV-2001; 2001WO-US043956.

XX 15-NOV-2000; 2000US-0249407P.

XX (INCY-) INCYTE GENOMICS INC.

XX Ramkumar J, Arvizu C;

XX WPI; 2002-537629/57.

XX N-PSDB; AAD41225.

XX New polypeptides of human embryogenesis associated proteins for screening

XX modulators useful for treating or preventing disorders e.g.

XX endometriosis, infertility, allergy, preclampsia.

XX Claim 57; Page 94-95; 97pp; English.

XX The invention relates to human embryogenesis associated proteins (EMERY)

XX and nucleic acid molecules encoding such proteins. EMERY sequences are

XX useful for screening modulators useful for treating or preventing

XX disorders associated with abnormal expression of EMERY. The disorders

XX treated include reproductive disorders such as infertility,

XX endometriosis, endometrial or ovarian tumour; autoimmune/inflammatory

XX disorder such as acquired immune deficiency syndrome (AIDS), allergies,

XX contact dermatitis; disorders of the placenta such as preclampsia, for

XX abruptio placentae etc. Sequences of the invention are also useful for

XX analysing a proteome of a tissue or a cell type. EMERY proteins are

XX useful as immunogens for preparing antibodies. Polynucleotides of the

XX invention are useful for creating knockin humanised animals or transgenic

XX animals to model human diseases. They are also used in gene therapy. The

XX present sequence is human EMERY-2 protein

XX SQ

Sequence 542 AA;

Query Match

Best Local Similarity 89.1%; Score 2558; DB 5; Length 542;

Matches 479; Conservative 22; Mismatches 37; Indels 4; Gaps 1;

480 EAVKQLQVLMQEPQWESMTKIHGRPLDRPASPCSDVNDIEGTPEEISTAQPLCPNSAG 539

Db 121 QGQAREKQVKEAISQITRGHSTSPYKGLVLSKLTETLTHRLVSLFNTTLRLHEVSA 180

Qy 181 QNPTNCWICLPLNFRPYVSIIPVPEOWNNEFTINTSVLGVPLVSNLEITHTSNLTVCVF 240

Db 181 QNPTNCWICLPLNFRPYVSIIPVPEOWNNEFTINTSVLGVPLVSNLEITHTSNLTVCVF 240

Qy 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300

Db 241 SNTIDTSSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300

Qy 301 TEODLYSVYISKPRKRVFIPVIGAGVGLGTCGIGITSTQFYKLSOELNGDMER 360

Db 301 TEODLYSVYISKPRKRVFIPVIGAGVGLGTCGIGITSTQFYKLSOELNGDMER 360

Qy 361 VADSLVTQDLQNSLAQVQLQNRALDLTAERGTCFLGEECCVYVNSQSVITVKVKE 420

Db 361 VADSLVTQDLQNSLAQVQLQNRALDLTAERGTCFLGEECCVYVNSQSVITVKVKE 420

Qy 421 IRDRIQRAEELNTGPGWGLSQMPEWILPFGPLAAIILLFGPCIFNLVNFVSSRI 480

Db 421 IRDRIQRAEELNTGPGWGLSQMPEWILPFGPLAAIILLFGPCIFNLVNFVSSRI 480

Qy 481 EAVK----LQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAG 536

Db 481 EAVKQLQVLMQEPQWESMTKIHGRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAG 540

Qy 537 SS 538

Db 541 SS 542

RESULT 15

AAW71068

ID AAW71068 standard; protein; 493 AA.

AC AAW71068;

XX AAW71068;

DT 29-DEC-1998 (first entry)

XX

DE Multiple sclerosis associated retrovirus protein 3.

XX

KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;

XX env gene; rheumatoid arthritis-associated virus.

XX

OS Multiple sclerosis associated retrovirus.

XX

PH Key Location/Qualifiers

FT Misc-difference 39 /note= "Encoded by tga"

XX

PN WO9823755-A1.

XX

PD 04-JUN-1998.

XX

PF 26-NOV-1997; 97WO-1B001482.

XX

PR 26-NOV-1996; 96US-00756429.

XX

PA (INMR) BIO MERIEUX.

XX

PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;

XX

DR WPI; 1998-322732/28.

XX

XX N-PSDB; AAV43217.

XX

PT New nucleic acid from retroviruses - useful for diagnosis, prevention and

XX treatment of, e.g. multiple sclerosis.

XX

PS Disclosure; Page 185-187; 286pp; English.

XX

CC The present sequence represents a multiple sclerosis (MS) associated

480 EAVKQLQVLMQEPQWESMTKIHGRPLDRPASPCSDVNDIEGTPEEISTAQPLCPNSAG 539

Db 121 QGQAREKQVKEAISQITRGHSTSPYKGLVLSKLTETLTHRLVSLFNTTLRLHEVSA 180

Qy 181 QNPTNCWICLPLNFRPYVSIIPVPEOWNNEFTINTSVLGVPLVSNLEITHTSNLTVCVF 240

Db 181 QNPTNCWICLPLNFRPYVSIIPVPEOWNNEFTINTSVLGVPLVSNLEITHTSNLTVCVF 240

Qy 241 SNTTYTNSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300

Db 241 SNTIDTSSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300

Qy 301 TEODLYSVYISKPRKRVFIPVIGAGVGLGTCGIGITSTQFYKLSOELNGDMER 360

Db 301 TEODLYSVYISKPRKRVFIPVIGAGVGLGTCGIGITSTQFYKLSOELNGDMER 360

Qy 361 VADSLVTQDLQNSLAQVQLQNRALDLTAERGTCFLGEECCVYVNSQSVITVKVKE 420

Db 361 VADSLVTQDLQNSLAQVQLQNRALDLTAERGTCFLGEECCVYVNSQSVITVKVKE 420

Qy 421 IRDRIQRAEELNTGPGWGLSQMPEWILPFGPLAAIILLFGPCIFNLVNFVSSRI 480

Db 421 IRDRIQRAEELNTGPGWGLSQMPEWILPFGPLAAIILLFGPCIFNLVNFVSSRI 480

Qy 481 EAVK----LQMEPKMQSKTKIYRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAG 536

Db 481 EAVKQLQVLMQEPQWESMTKIHGRPLDRPASPRSDVNDIKGTPPEISAAQPLLRNSAG 540

Qy 537 SS 538

Db 541 SS 542

RESULT 15

AAW71068

ID AAW71068 standard; protein; 493 AA.

AC AAW71068;

XX AAW71068;

DT 29-DEC-1998 (first entry)

XX

DE Multiple sclerosis associated retrovirus protein 3.

XX

KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene; gag gene;

XX env gene; rheumatoid arthritis-associated virus.

XX

OS Multiple sclerosis associated retrovirus.

XX

PH Key Location/Qualifiers

FT Misc-difference 39 /note= "Encoded by tga"

XX

PN WO9823755-A1.

XX

PD 04-JUN-1998.

XX

PF 26-NOV-1997; 97WO-1B001482.

XX

PR 26-NOV-1996; 96US-00756429.

XX

PA (INMR) BIO MERIEUX.

XX

PI Perron H, Beseme F, Bedin F, Paranhos-Baccala G;

PI Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B;

XX

DR WPI; 1998-322732/28.

XX

XX N-PSDB; AAV43217.

XX

PT New nucleic acid from retroviruses - useful for diagnosis, prevention and

XX treatment of, e.g. multiple sclerosis.

XX

PS Disclosure; Page 185-187; 286pp; English.

XX

CC The present sequence represents a multiple sclerosis (MS) associated

CC retrovirus (MSRV) protein used in the method of the invention. The
 CC invention provides complete or partial genomic sequences of the MSRV-1
 CC pol gene, gag gene and env gene, and polypeptides encoded by these genes.
 CC The invention also provides antibodies raised against the polypeptides.
 CC The genomic sequences, polypeptides and antibodies are also claimed
 CC useful for diagnosing infection by MS and rheumatoid arthritis-associated
 CC viruses, and also for prevention and treatment of infection with these
 CC viruses
 CC
 SQ Sequence 493 AA;

Query Match 80.6%; Score 2314; DB 2; Length 493;
 Best Local Similarity 88.4%; Pred. No. 1e-205;
 Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLEMRQPGNIDAPSYRSLSKGTP 60
 Db 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLEMRQPGNIDAPSYRSLSKGNS 60

Qy 61 TFTAHTMPNCVHSATLQMHANTHYWTKMINPSCPGGLGVTVCTYFTQIGMSDGGV 120
 Db 61 TFTAHTMPNCVNSATLQMHANTHYWTKMINPSCPGGLGATVCTYFTHTSMSDGGGI 120

Qy 121 QDOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180
 Db 121 QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180

Qy 181 QNPNCWICPLNRPVYSIPVPEQWNNFSTEINTSVLVGPLVSNLEITHSNLTCVKF 240
 Db 181 QNPNCWMCPLNRPVYSIPVPEQWNNFSTEINTSVLVGPLVSNLEITHSNLTCVKF 240

Qy 241 SNTTYTNSQICIRWTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300
 Db 241 SNTIDTSSQICIRWTPPTQIVCLPSGIFVCGTSAYHCLNGSSSMCFSLFVPPMTIY 300

Qy 301 TEQDLYSYVSKPRNKRVPILPFFVIGAGVLGALGTGIGTITTSQFYFKLSQELNGDMER 360
 Db 301 TEQDLYNHVVPKPNKRVPILPFFVIRAGVLGRLGTGIGTITTSQFYFKLSQELNGDMEQ 360

Qy 361 VADSLVTLQDOLNSLAVALQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 Db 361 VTSLSVTLQDOLNSLAVALQNRALDILLTAERGCTCLFLGEECCYVYVNSQRIUTEKVK 420

Qy 421 IRDRIQRAEELNTPGWLSSQMPWLPFLGFLAAAILLLFGPCIFNLLVNFVSSRI 480
 Db 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGFLAAAILLLFGPCIFNLLVNFVSSRI 480

Qy 481 EAVKLQMEPKMQ 492
 Db 481 EAVKLQWLQME 492

RESULT 16
 AA99553
 ID AA99553 standard; protein; 493 AA.
 AC AA99553;
 DT 17-OCT-2003 (revised)
 DT 27-AUG-2003 (revised)
 DT 08-JUN-1999 (first entry)
 DE Protein encoded by clone C15 from MSRV-1.
 KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KW rheumatoid polyarthritis.
 OS Viruses.
 PH Key Location/Qualifiers
 FT Misc-difference 39 /label= unknown
 FT
 XX

PN FR2765588-A1.
 XX 08-JAN-1993.
 XX 07-JUL-1997; 97FR-00008816.
 XX 07-JUL-1997; 97FR-00008816.
 XX (INNR) BIO MERIEUX.
 XX WPI; 1999-098275/09.
 XX N-PSDB; AAX29703.
 DR Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 XX multiple sclerosis or rheumatoid polyarthritis.
 CC Claim 2; Page 38-39; 83pp; French.
 CC This sequence represents the protein encoded by clone C15 from a novel
 CC multiple sclerosis related virus type 1 (MSRV1). The sequence can be used
 CC in diagnostic, prophylactic or therapeutic compositions to inhibit
 CC expression of a multiple sclerosis related virus and/or virus associated
 CC with rheumatoid polyarthritis. (Updated on 27-AUG-2003 to correct OS
 CC field.) (Updated on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 493 AA;

Query Match 80.6%; Score 2314; DB 2; Length 493;
 Best Local Similarity 88.4%; Pred. No. 1e-205;
 Matches 435; Conservative 19; Mismatches 38; Indels 0; Gaps 0;

Qy 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLEMRQPGNIDAPSYRSLSKGTP 60
 Db 1 MALPYHIFLFTVLLPSFTLTAPPCCRCMTSSSPYQEFLEMRQPGNIDAPSYRSLSKGNS 60

Qy 61 TFTAHTMPNCVHSATLQMHANTHYWTKMINPSCPGGLGVTVCTYFTQIGMSDGGV 120
 Db 61 TFTAHTMPNCVNSATLQMHANTHYWTKMINPSCPGGLGATVCTYFTHTSMSDGGGI 120

Qy 121 QDOAREKHVKEVISQLTRVHGTSPPYKGLDLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180
 Db 121 QGOAREKQVKEAISQLTRGHSTSPYKGLVLSKLHETLRTHTRLVSLFNTLTGLHEVSA 180

Qy 181 QNPNCWICPLNRPVYSIPVPEQWNNFSTEINTSVLVGPLVSNLEITHSNLTCVKF 240
 Db 181 QNPNCWMCPLNRPVYSIPVPEQWNNFSTEINTSVLVGPLVSNLEITHSNLTCVKF 240

Qy 241 SNTTYTNSQICIRWTPPTQIVCLPSGIFVCGTSAYRCINGSSSMCFSLFVPPMTIY 300
 Db 241 SNTIDTSSQICIRWTPPTQIVCLPSGIFVCGTSAYHCLNGSSSMCFSLFVPPMTIY 300

Qy 301 TEQDLYSYVSKPRNKRVPILPFFVIGAGVLGALGTGIGTITTSQFYFKLSQELNGDMER 360
 Db 301 TEQDLYNHVVPKPNKRVPILPFFVIRAGVLGRLGTGIGTITTSQFYFKLSQELNGDMEQ 360

Qy 361 VADSLVTLQDOLNSLAVALQNRALDILLTAERGCTCLFLGEECCYVYVNSGIVTEKVK 420
 Db 361 VTSLSVTLQDOLNSLAVALQNRALDILLTAERGCTCLFLGEECCYVYVNSQRIUTEKVK 420

Qy 421 IRDRIQRAEELNTPGWLSSQMPWLPFLGFLAAAILLLFGPCIFNLLVNFVSSRI 480
 Db 421 IRDRIQRAEELQNTERWGLLSQMPWLPFLGFLAAAILLLFGPCIFNLLVNFVSSRI 480

Qy 481 EAVKLQMEPKMQ 492
 Db 481 EAVKLQWLQME 492

RESULT 17
 ADB84401
 ID ADB84401 standard; protein; 493 AA.
 XX
 AC ADB84401;

XX 04-DEC-2003 (first entry)
 XX MSRV-1 associated protein/peptide sequence #10.
 XX multiple sclerosis; rheumatoid arthritis; gag; pol;
 XX reverse transcriptase; ribonuclease H.
 XX Unidentified;
 XX US2003039664-A1.
 XX 27-FEB-2003.
 XX 26-NOV-1997; 97US-00979847.
 XX 26-NOV-1996; 96US-00756429.
 XX (PERR/) PERRON H.
 XX (BESE/) BESEME F.
 XX (BEDI/) BEDIN F.
 XX (PARA/) PARANHOS-BACCALA G.
 XX (KOMU/) KOMURIAN-PRADEL F.
 XX (JOLI/) JOLIVET-REYNAUD C.
 XX (MAND/) MANDRAND B.
 XX (GARS/) GARSON J A.
 XX (TUKES/) TUKES P W.
 XX Perron H, Beseme F, Bedin F, Paranhos-Baccala G;
 XX Komurian-Pradel F, Jolivet-Reynaud C, Mandrand B, Garson JA, Tuke PW;
 XX WPI; 2003-51253/48.
 XX New isolated or purified nucleic acid associated with multiple sclerosis
 XX and/or rheumatoid arthritis, useful for detecting a virus associated with
 XX multiple sclerosis or rheumatoid arthritis in a biological sample.
 XX Claim 31; Page 78-80; 193pp; English.
 XX The invention relates to an isolated or purified nucleic acid from a
 XX virus associated with multiple sclerosis and/or rheumatoid arthritis,
 XX multiple sclerosis-associated virus (MSRV)-1. The nucleic acids comprise
 XX pol, gag or reverse transcriptase genes (or their fragments) encoding the
 XX proteins or defined peptides (including immunodominant peptides,
 XX antigenic peptides or conserved motifs). Also included are a process for
 XX detecting a virus associated with multiple sclerosis or rheumatoid
 XX arthritis in a biological sample, a nucleic acid probe for the detection
 XX of a virus associated with multiple sclerosis or rheumatoid arthritis, a
 XX primer for the amplification by polymerisation of a nucleic acid of a
 XX viral material associated with multiple sclerosis or rheumatoid
 XX arthritis, a polypeptide exhibiting an inhibitory activity on the
 XX proteolytic, reverse transcriptase or ribonuclease H activity from MSRV,
 XX and an antibody directed against the MSRV-1 virus obtained by an
 XX immunologically reacting a human or animal body or cells with an
 XX immunogenic agent consisting of the antigenic polypeptide defined above.
 XX The nucleic acids are useful for detecting a biological sample a virus
 XX associated with multiple sclerosis or rheumatoid arthritis, or for
 XX detecting in a biological sample, the presence of or exposure to a virus
 XX associated with multiple sclerosis or rheumatoid arthritis. The present
 XX sequence is a claimed MSRV-associated sequence whose identity cannot be
 XX accurately determined. Note: The SEQ ID numbers for the sequences as
 XX displayed in the main body of the patent do not match the SEQ ID numbers
 XX in the sequence listing. Consequently these sequences mentioned in the
 XX claims may not be the sequences the authors intended to claim.
 XX Sequence 493 Aa;
 XX Query Match 79.7%; Score 2289; DB 7; Length 493;
 XX Best Local Similarity 87.4%; Pred. No. 2.2e-203;
 XX Matches 430; Conservative 21; Mismatches 41; Indels 0; Gaps 0;
 XX 1 MALPHYHFLFTVLLPPFALTAPPCCCTSSSPYQBFLLXNRLPGNIDAPSYSLSKGNS 60
 XX
 XX DB 1 MALPHYHFLFTVLLPPFALTAPPCCCTSSSPYQBFLLXNRLPGNIDAPSYSLSKGNS 60
 XX QY 61 TFTAHTMPNRCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVVCWTFYTTQTKMSDGGV 120
 XX DB 61 TFTAHTMPNRCYHSATLCHMANTHYWTGKMINPSCPGGLGVTVVCWTFYTTQTKMSDGGV 120
 XX QY 121 QDQAREKHVEVISQLTRVHGTSPYKGLDL SKLHETLRHTRLVSLFNTLTLGLHEVSA 180
 XX DB 121 QDQAREKHVEVISQLTRVHGTSPYKGLDL SKLHETLRHTRLVSLFNTLTLGLHEVSA 180
 XX QY 181 QNPTNCWICLPLNFRPYVSIPEQNNFSTEINTTSVLVGPLVSNLEIHTSNLTCVKF 240
 XX DB 181 QNPTNCWICLPLNFRPYVSIPEQNNFSTEINTTSVLVGPLVSNLEIHTSNLTCVKF 240
 XX QY 241 SNTYTTNSOCIRWVTPPTQIVCLPSGIFPVCGTSAYRCUNGSESMCFSLFVPPMTIY 300
 XX DB 241 SNTYTTNSOCIRWVTPPTQIVCLPSGIFPVCGTSAYRCUNGSESMCFSLFVPPMTIY 300
 XX QY 301 TEQDLXSVISKPRNKRPVILPFVIGAGVLGALGTGGITTTSTQFYFKLSQELNGDMER 360
 XX DB 301 TEQDLXSVISKPRNKRPVILPFVIGAGVLGALGTGGITTTSTQFYFKLSQELNGDMER 360
 XX QY 361 VADSLVTLOQLNSLAAVLQNRRLDLPFAERGCTCLFLGEECCYVYVNGSIVTEKVK 420
 XX DB 361 VADSLVTLOQLNSLAAVLQNRRLDLPFAERGCTCLFLGEECCYVYVNGSIVTEKVK 420
 XX QY 421 IRDRIQRABELNTPGWLISQWMPILPFLGLAAILLLLPFCIEFLLVNFVSSRI 480
 XX DB 421 IRDRIQRABELNTPGWLISQWMPILPFLGLAAILLLLPFCIEFLLVNFVSSRI 480
 XX QY 481 EAVKLQWPEKMQ 492
 XX DB 481 EAVKLQWPEKMQ 492
 XX
 XX RESULT 18
 XX AAM18008
 XX ID AAM18008 standard; protein; 263 AA.
 XX AC AAM18008;
 XX DT 12-OCT-2001 (first entry)
 XX DE Peptide #4442 encoded by probe for measuring cervical gene expression.
 XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
 XX KW cervical cancer.
 XX OS Homo sapiens.
 XX PN WO200157278-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US000670.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX FA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488901/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human cervical epithelial cells.

Claim 27; SEQ ID NO 22834; 487bp; English.

The present invention relates to human single exon nucleic acid probes (SENPs; see AA10068-AA128459). The present sequence is a peptide encoded by one such probe. The SENPs are derived from human Hela cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 263 AA;

Query Match 43.0%; Score 1234.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 8.2e-106;
Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;

QY 126 EKHVKEVISQTRVHGTSPPYKGLDLSKLEHETLTHTRLVSLFNTLTGLHEVSAQNPTN 185
DB 1 EKHVKEVISQTRVHGTSPPYKGLDLSKLEHETLTHTRLVSLFNTLTGLHEVSAQNPTN 60
QY 186 CWICLPLNFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTTY 245
DB 61 CWICLPLNFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTTY 119
QY 246 TTNQCIRWVTPPTQIVCLPSGIFVCGTSAYRCNGSSSMCFSLFLVPPMTIYTEQDL 305
DB 120 TANSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCNGSSSMCFSLFLVPPMTIYTEQDL 179
QY 306 YSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDMERVADSL 365
DB 180 YNHVIPKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDMERVADSL 239
QY 366 VTLDQDLNSLAADVQLNRRALDILL 389
DB 240 VTLDQDLNSLAADVQLNRRALDILL 263

RESULT 19
ABB37043
ID ABB37043 standard; peptide; 263 AA.
AC ABB37043;
XX
DT 04-FEB-2002 (first entry)
DE Peptide #4549 encoded by human foetal liver single exon probe.
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human foetal liver.

XX PS Claim 27; SEQ ID NO 29678; 639pp + Sequence Listing; English.

XX CC The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human foetal liver. The present invention is a peptide encoded by a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 263 AA;

Query Match 43.0%; Score 1234.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 8.2e-106;
Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;

QY 126 EKHVKEVISQTRVHGTSPPYKGLDLSKLEHETLTHTRLVSLFNTLTGLHEVSAQNPTN 185
DB 1 EKHVKEVISQTRVHGTSPPYKGLDLSKLEHETLTHTRLVSLFNTLTGLHEVSAQNPTN 60
QY 186 CWICLPLNFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTTY 245
DB 61 CWICLPLNFRPYISIPVPEQWNNFSTEINTSVLVGPLVSNLEITHTSNLTVCVKSNTTY 119
QY 246 TTNQCIRWVTPPTQIVCLPSGIFVCGTSAYRCNGSSSMCFSLFLVPPMTIYTEQDL 305
DB 120 TANSQCIRWVTPPTQIVCLPSGIFVCGTSAYRCNGSSSMCFSLFLVPPMTIYTEQDL 179
QY 306 YSYVVISKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDMERVADSL 365
DB 180 YNHVIPKPRNKRVPILPFVIGAGVLGALGTGIGGITTSTQFYKLSQELNGDMERVADSL 239
QY 366 VTLDQDLNSLAADVQLNRRALDILL 389
DB 240 VTLDQDLNSLAADVQLNRRALDILL 263

RESULT 20
AAM30519
ID AAM30519 standard; protein; 263 AA.
AC AAM30519;
XX
DT 17-OCT-2001 (first entry)
DE Peptide #4556 encoded by probe for measuring placental gene expression.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX Homo sapiens.
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 PT
 XX Claim 27; SEQ ID NO 30788; 654pp; English.
 XX
 XX The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders
 XX
 XX Sequence 263 AA;
 SQ
 Query Match 43.0%; Score 1234.5; DB 4; Length 263;
 Best Local Similarity 89.4%; Pred. No. 8.2e-106;
 Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;
 QY 126 EKHVEVISQLTRVHGTSPPYKGLDLSKLHETLRTHRLVSLFNTLTGLHEVSAQNPTN 185
 DB 1 EKHKEVISQLTWVHSTPGPYKGLDLSKLHETLRTHRLVSLFNTLTGLHEVSAQNPTN 60
 QY 186 CWICLPLNFRPVYSIPVPEQWNNFSTEINTTSVLGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 245
 DB 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 119
 QY 246 TTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVYCLNGSSVMCFSLFVPPMTIYTEQDL 305
 DB 120 TANSQCIRWVTPPTQIVCLPSGIFVCGTSAVYCLNGSSVMCFSLFVPPMTIYTEQDL 179
 QY 306 YSVVISKPRNKRVPILPFVIGAGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 365
 DB 180 YNHVIPKPRNKRVPILPFVIGAGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 239
 QY 366 VTLODQLNSLAADVVLQNRALDILL 389
 DB 240 VTLODQLNSLAADVVLQNRALDILL 263
 RESULT 21
 ABB31812
 ID ABB31812 standard; peptide; 263 AA.
 XX
 AC ABB31812;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Peptide #4463 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human; microarray; single exon probe; gene expression; breast; disease;
 KW cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US0000662.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-496933/54.
 XX New spatially-addressable set of single exon nucleic acid probes, useful
 PT for measuring gene expression in sample derived from human breast,
 PT comprises number of single exon nucleic acid probes.
 XX
 XX Claim 27; SEQ ID NO 14780; 327pp + Sequence Listing; English.
 XX
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting the
 CC probes with a collection of detectably labelled nucleic acids derived
 CC from mRNA of human breast, and then measuring the label bound to each
 CC probe of the microarray. The probes are useful for verifying the
 CC expression of regions of genomic DNA predicted to encode proteins. They
 CC are useful for gene discovery, and for determining predisposition and/or
 CC prognosing breast disease. Gene expression analysis is useful for
 CC assessing the toxicity of chemical agents on cells. The microarray of
 CC this invention presents a far greater diversity of probes for measuring
 CC gene expression, with far less bias than expressed sequence tag
 CC microarrays. The method is suitable for rapid production of functional
 CC information from genomic sequence. The present sequence is a peptide
 CC encoded by a single exon nucleic acid probe of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 263 AA;
 SQ
 Query Match 43.0%; Score 1234.5; DB 4; Length 263;
 Best Local Similarity 89.4%; Pred. No. 8.2e-106;
 Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;
 QY 126 EKHVEVISQLTRVHGTSPPYKGLDLSKLHETLRTHRLVSLFNTLTGLHEVSAQNPTN 185
 DB 1 EKHKEVISQLTWVHSTPGPYKGLDLSKLHETLRTHRLVSLFNTLTGLHEVSAQNPTN 60
 QY 186 CWICLPLNFRPVYSIPVPEQWNNFSTEINTTSVLGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 245
 DB 61 CWMCLPLHFRPYISIPVPEQWNNFSTEINTTSVLGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 119
 QY 246 TTNSQCIRWVTPPTQIVCLPSGIFVCGTSAVYCLNGSSVMCFSLFVPPMTIYTEQDL 305
 DB 120 TANSQCIRWVTPPTQIVCLPSGIFVCGTSAVYCLNGSSVMCFSLFVPPMTIYTEQDL 179
 QY 306 YSVVISKPRNKRVPILPFVIGAGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 365
 DB 180 YNHVIPKPRNKRVPILPFVIGAGVLGALGTGGITTTQFYKLSQELNGDMERVADSL 239
 QY 366 VTLODQLNSLAADVVLQNRALDILL 389
 DB 240 VTLODQLNSLAADVVLQNRALDILL 263
 RESULT 22
 ABB22358
 ID ABB22358 standard; protein; 263 AA.
 XX
 AC ABB22358;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Protein #4357 encoded by probe for measuring heart cell gene expression.
 XX
 KW Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 XX
 OS Homo sapiens.
 XX


```
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29875.
XX DE
XX DE Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024253.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX DR
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains.
XX PS Example 4; SEQ ID NO 29875; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention
XX SQ . Sequence 263 AA;
Query Match 43.0%; Score 1234.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 8.2e-106;
Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;
QY 126 EKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTLTVSLVLENTLTGLHEVSAQNPTN 185
DB 1 EKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTLTVSLVLENTLTGLHEVSAQNPTN 60
QY 186 CWICLPNFRPYVSIPIVPEQWNNFSTEINTTSLVGLVLSNLEITHTSNLTVCVKSNTTY 245
DB 61 CWICLPNFRPYVSIPIVPEQWNNFSTEINTTSLVGLVLSNLEITHTSNLTVCVKSNTTY 119
QY 246 TTNSQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSVCFLSFLVAPMPITYEODL 305
DB 120 TANSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSVCFLSFLVAPMPITYEODL 179
QY 306 YSVVSKPRNKRVPILPFVIGAGVLCALGTGIGGITTSTQFYKLSQELNGDMERVADSL 365
DB 180 YNHVLPKRNKRVPILPFVIGAGVLCALGTGIGGITTSTQFYKLSQELNGDMERVADSL 239
QY 366 VTLDQQLNSLAAVLQNRALDIL 389
DB 240 VTLDQQLNSLAAVLQNRALDIL 263
RESULT 25.
ABG51885
ID ABG51885 standard; peptide; 263 AA.
XX
AC ABG51885;
```

```
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 30533.
XX DE
XX DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX OS
XX OS Homo sapiens.
XX PN WO200157273-A2.
XX PD
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 30533; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.
XX CC (I) may be used for predicting, measuring and displaying gene expression
XX CC in samples derived from human adult liver. The genes identified may be
XX CC involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX CC associated with coronary heart disease. ABG47348-ABG59330 represent human
XX CC liver single exon encoded peptides of the invention. Note: The sequence
XX CC information for this patent does not appear in the printed specification
XX CC but was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ . Sequence 263 AA;
Query Match 43.0%; Score 1234.5; DB 4; Length 263;
Best Local Similarity 89.4%; Pred. No. 8.2e-106;
Matches 236; Conservative 11; Mismatches 16; Indels 1; Gaps 1;
QY 126 EKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTLTVSLVLENTLTGLHEVSAQNPTN 185
DB 1 EKHVKEVISQLTRVHGTSSPYKGLDLSKLHETLRTHTLTVSLVLENTLTGLHEVSAQNPTN 60
QY 186 CWICLPNFRPYVSIPIVPEQWNNFSTEINTTSLVGLVLSNLEITHTSNLTVCVKSNTTY 245
DB 61 CWICLPNFRPYVSIPIVPEQWNNFSTEINTTSLVGLVLSNLEITHTSNLTVCVKSNTTY 119
QY 246 TTNSQICIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSVCFLSFLVAPMPITYEODL 305
DB 120 TANSQCIRWTPPTQIVCLPSGIFVCGTSAYRCLNGSSSVCFLSFLVAPMPITYEODL 179
QY 306 YSVVSKPRNKRVPILPFVIGAGVLCALGTGIGGITTSTQFYKLSQELNGDMERVADSL 365
DB 180 YNHVLPKRNKRVPILPFVIGAGVLCALGTGIGGITTSTQFYKLSQELNGDMERVADSL 239
QY 366 VTLDQQLNSLAAVLQNRALDIL 389
```


Db 240 VTQDQLNSLAAYVLQNRALDLL 363

Search completed: April 19, 2004, 10:36:49
Job time : 65 secs